Bank of Canada Banque du Canada

Working Paper 2001-15 / Document de travail 2001-15

Affine Term-Structure Models: Theory and Implementation

by

David Jamieson Bolder

ISSN 1192-5434

Printed in Canada on recycled paper

Bank of Canada Working Paper 2001-15

October 2001

Affine Term-Structure Models: Theory and Implementation

by

David Jamieson Bolder

Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
dbolder@bankofcanada.ca

The views expressed in this paper are those of the author.
No responsibility for them should be attributed to the Bank of Canada.

101-07621/

Contents

Acknowledgements ..... **iv**

Abstract/Résumé ..... **v**

1 Introduction **1**

2 Theoretical Development **5**

2.1 Some key interest rate relationships 6

2.2 The basic structure 7

2.3 The single-factor model 9

2.4 The multifactor model 18

3 Model Implementation **27**

3.1 The Kalman filter in brief 28

3.2 The state-space formulation 28

3.3 The Kalman filter in detail 31

3.4 Simulation results 34

3.5 Actual results 38

4 Conclusion **41**

Appendixes **42**

A The One-Factor CIR Model 42

B Solving the SDE 47

C Discrete-Time Affine Models 50

Bibliography **56**

Acknowledgements

I would like to especially thank Toni Gravelle and Mark Reesor of the Bank of Canada, Phelim Boyle and Adam Kolkiewicz of the University of Waterloo, James Morley of Washington University, and Jeremy Graveline of Stanford University for fruitful discussions, timely assistance, and diligently reading preliminary versions of this paper. I also would like to thank Peter Thurlow, Walter Engert, Ron Morrow, André Bernier, Chris D'Souza, and Ben Fung of the Bank of Canada for their assistance. Any errors, omissions, or inconsistencies are, of course, my sole responsibility.

Abstract

Affine models describe the stylized time-series properties of the term structure of interest rates in a reasonable manner, they generalize relatively easily to higher dimensions, and a vast academic literature exists relating to their implementation. This combination of characteristics makes the affine class a natural introductory point for modelling interest rate dynamics. The author summarizes and synthesizes the theoretical and practical specifics relating to this analytically attractive class of models. This summary is accomplished in a self-contained manner with sufficient detail so that relatively few technical points will be left for the reader to ponder. As such, this paper represents a first step towards advancing the Bank of Canada's research agenda in this area, with a view to using these models to assist with practical debt and risk-management problems currently under study.

JEL classification: C0, C5, G0

Bank classification: Interest rates; Econometric and statistical methods; Debt management

Résumé

Il existe actuellement une abondante littérature sur les modèles affines, qui décrivent assez bien les propriétés en séries temporelles de la structure à terme des taux d'intérêt et permettent, d'une façon relativement simple, de généraliser les équations multidimensionnelles. Ces caractéristiques des modèles affines en font un bon point de départ la modélisation des propriétés dynamiques des taux d'intérêt. La présente étude a pour but de présenter une synthèse, à la fois théorique et pratique, des modèles affines, fort attrayants du point de vue analytique. L'auteur traite en profondeur le sujet est laisse peu de points techniques en suspens. Avec cette publication, la Banque fait un premier pas vers l'utilisation de ces modèles aux fins règlement des problèmes pratiques sur lesquels elle se penche présentement dans les domaines de la gestion du risque et de la dette.

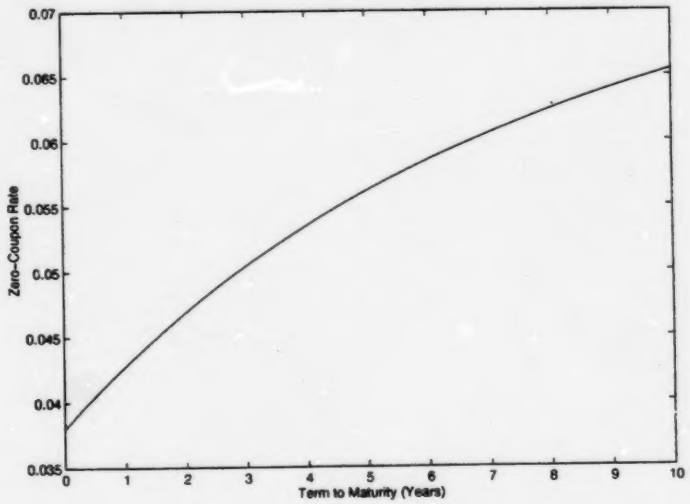
Classification JEL : C0, C5, G0

Classification de la Banque : Taux d'intérêt; Méthodes économétriques et statistiques; Gestion de la dette

1 Introduction

Term-structure modelling refers to one of two distinct, albeit related, problems in finance. The first problem involves fitting a zero-coupon interest rate curve to a set of cross-sectional bond price observations. Figure 1 displays this familiar relationship between the zero-coupon interest rate and its term to maturity. This relationship is generally called the *term structure of interest rates*. Loosely speaking, the solution to this first problem is equivalent to taking a "snapshot" of the term structure of interest rates at a given instant in time.[1]

Figure 1: **The Term Structure at an Instant in Time**: The following graphic outlines the relationship between zero-coupon rates and their term to maturity at a given instant in time.

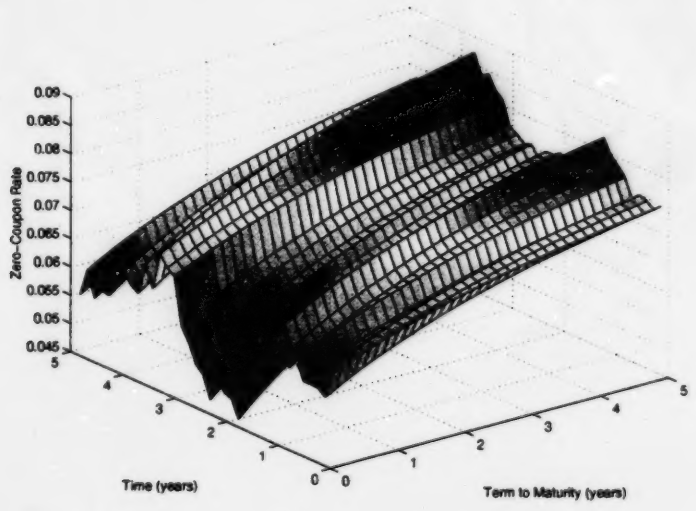


The second problem, which is the focus of this paper, relates to the specification of the intertemporal dynamics of the entire term structure of interest rates. Simply put, it poses the question: *"how does the term structure of interest rates evolve over time?"* This is conceptually similar to asking how a stock price or an exchange rate moves through time. In finance and economics we are accustomed to thinking of key variables as evolving over time in some unpredictable fashion; they are, in a technical and figurative sense, random variables. The central theoretical and empirical fact about stock prices, exchange rates, economic growth rates, and even inflation is that they cannot be forecast with certainty. From a modelling perspective, therefore, we typically represent these random variables as scalar-valued stochastic processes exhibiting

[1] For a detailed overview of this issue, see Bolder and Stréliski (1999).

varying degrees of volatility, boundedness, and mean-reversion. Interest rate modelling, however, is a more complicated undertaking, because the term structure is not a scalar quantity—such as the exchange rate or a particular commodity price—but is actually a vector-valued object. Thus, as compared to the first problem, the addition of a time dimension significantly increases the complexity of the problem. Not only must we ensure that the dynamics of a given term-structure model are reasonable from an empirical perspective, but we also require that the model be free of arbitrage. That is, the movements of the term structure do not permit conditions to occur under which market participants may earn risk-free profits. Henceforth in this paper, the expression *term-structure modelling* will refer to this second problem. Figure 2 illustrates a sample path from one of the models considered in detail later in the paper. Observe how the term structure of interest rates moves over time. At each instant, we can slice across the time axis and recover the cross-sectional term structure of interest rates, as described in Figure 1.

Figure 2: **Term-Structure Dynamics over Time**: This graphic details the evolution of the term structure of interest rates (as summarized in Figure 1) over a five-year time horizon.



The previous definition begs the question *"why do we need to be able to solve this problem?"* The answer lies in the fact that we can use these models to solve real problems. The primary application of term-structure models in the finance literature is towards the determination of prices for interest-rate-contingent claims.[2] While this is an important usage of these models, their immediate application at the Bank of

[2]A contingent claim is a financial claim that pays its cash-flow, or cash-flows, only if some predetermined state of the world ($\omega \in \Omega$) is achieved at some point, or points, in the future. Some common examples of interest-rate-contingent claims include

Canada is as an input into debt strategy and risk-management analysis. Specifically, *debt strategy analysis* deals with determining the appropriate issuance pattern for Government of Canada debt. For example, one central question is "how much of the federal government's annual borrowing needs should be financed with long-term coupon bonds versus short-term treasury bills?" In this context, it is useful to consider how various issuance strategies perform under different interest rate outcomes. This type of analysis requires a methodology for generating future interest rate scenarios. In fact, a prerequisite for solving this problem is an understanding of term-structure dynamics, because future term-structure outcomes are integral in the simulation of distributions for government debt charges that assist in deciding among various financing strategies. In a risk-management setting, the set of applications is similar. For example, the *value at risk* (VaR) of a fixed-income portfolio can be computed numerically by simulating paths for the term structure of interest rates and generating a distribution of values for the underlying portfolio under each scenario.[3] Therefore, to perform debt strategy and risk-management analysis, it is necessary to be able to adequately model the stochastic nature of the term structure of interest rates.

Having established that important applications exist for these models, the natural question then follows *"which model should we use?"* This question becomes critical when we consider that the academic and practitioner literature in this area is vast.[4] To address this issue, therefore, it is useful to consider what we are looking for in a model. This highlights the important point that model selection and the ultimate application are inextricably linked. There are three desired characteristics in a term-structure model for use in Bank of Canada applications:

- A model that adequately captures the dynamics of the Canadian term structure. If the model is not realistic, then it will not be particularly useful in a simulation exercise.
- The existence of an analytic representation for the relationship between the factors, or state variables, and the term structure of interest rates. Many models do not have this property. This is critical in a Monte Carlo setting where the speed of computation is often a key element of a viable algorithm.
- A parameter set that is relatively easy to estimate and interpret.

These considerations point us towards a rather rich set of term-structure models, popularized by Duffie and Kan (1996), termed the class of *affine* term-structure models. This class encompasses the models of Vasicek (1977), Cox, Ingersoll, and Ross (1985a,b), Longstaff and Schwartz (1992a,b), and a number of others. These models are formulated by assuming that future dynamics of the term structure of interest rates depend on the evolution of some observed, or unobserved, factor. This factor, also termed a state variable, is

caps, floors, swaptions, and callable bonds.

[3]Value at risk is defined as the maximum expected decrease in the value of a portfolio with a predefined probability over a predefined time horizon.

[4]An excellent overview of the issues involved in selecting a term-structure model is given by Rogers (1995).

a random process that is restricted by the assumption of an absence of arbitrage in the underlying financial market. The no-arbitrage restriction permits the derivation of a deterministic relationship between the term structure of interest rates and these state variables. Two special cases of this model, the Cox, Ingersoll, and Ross (CIR) and the Vasicek model, can readily be extended to incorporate multiple-state variables and permit analytic solutions for the bond price function; these two models will be considered in detail. This is important because substantial evidence suggests that the use of a single-state variable, or factor, to explain the random future movement of the term structure is inadequate. This inadequacy stems from the fact that the dynamics of the term structure of interest rates are too complex to be summarized by a single source of uncertainty. Empirical papers that explore this issue in detail include Chan, Karolyi, Longstaff, and Sanders (1992), Canabarro (1995), and Mc Manus and Watt (1999).

There are, nonetheless, some criticisms of this class of models. The first is that it is rare, even given the optimal parameter set, that these models will be able to adequately fit the current observed term structure of interest rates. In other words, affine models focus on the time-series properties of the term structure rather than its initial cross-sectional properties. A variety of models exist, including those suggested by Ho and Lee (1986) and Hull and White (1990, 1993, 1994a,b), that use numerical techniques to fit the model's initial term structure to the observed current term structure. An unfortunate side effect of the fitted models is that they have difficulty in adequately describing term-structure dynamics over long periods of time. This close fit to the current term structure is of critical importance to the pricing of contingent claims. It is, however, less important for the Bank of Canada's desired application, which mainly focuses on the simulation of term-structure sample paths. Indeed, for the purposes of simulation, the essential issue is that the time-series properties of the model adequately represent those of the Canadian term structure.

A second concern with the affine class of models is that they are inherently linear. Affine term-structure models are constructed by assuming that the bond price function is a linear function of the underlying state variables that provide the uncertainty in the model.[5] To address this issue, a number of models incorporate jumps into the diffusion processes that describe the dynamics of our state variables. While incorporation of non-linearities into the model is clearly a substantial increase in reality, it nonetheless represents a substantial increase in complexity. In particular, it often precludes the existence of closed-form solutions for the bond price function.

Acknowledging the validity of these criticisms, the decision to focus on the affine class of models primarily results from the fact that these models have the minimal set of characteristics—in terms of analytic tractability and time-series properties—to meet our desired application. Moreover, the affine class of term-structure models is in many respects the introductory point for such modelling. As a consequence, any subsequent work in this area will require a firm understanding of the fundamentals of the set of affine term-structure

[5]The theoretical development of these models is considered in section 2.

models. This is, therefore, a necessary first step in advancing the Bank of Canada's research agenda in this area.

The remainder of this document is intended primarily, as the title suggests, as an introduction to the theory and implementation of the class of affine term-structure models. The first goal is to provide a self-contained overview of the concepts required to understand this branch of the term-structure literature. The second goal is to consider a statistical method that can be used to estimate the parameter set of each model. Parameter estimation is the key issue for any practical implementation of the affine class of term-structure models. To this end, the paper is divided into two sections: theoretical development and model implementation. The first part, section 2, details the theoretical development of the models in continuous time. It begins with the single-factor model and then generalizes these concepts into a multidimensional setting. The second part of the paper, section 3, considers one possible technique for the estimation of the model parameters. Parameter estimation, for this class of models, is essentially a time-series problem. This is actually a rather difficult task as the state variables that govern the dynamics of the term structure are unobservable in this setting. The approach used requires that the models be represented in state-space form and uses a technique termed Kalman filtering to estimate the parameter set. This technique is extremely helpful when specifying parameters relating to unobserved state variables. To assess the accuracy of this estimation technique, the parameters are estimated using both simulated and actual zero-coupon interest rate data. It is important to note, however, that this is not an empirical paper and the estimation results are both exploratory and illustrative.

2 Theoretical Development

Often, the entry point for any modelling discussion is the discrete-time formulation. Nevertheless, in this paper we restrict our focus to the continuous-time formulation of the affine class of models.[6] Moreover, the discussion of model implementation in section 3 focuses exclusively on continuous models. Continuous-time finance models are conceptually demanding. They require some rather subtle results from the stochastic calculus and the theory of partial differential equations. The incremental effort, however, is generally worthwhile in terms of more complete models that are better able to solve important problems. Moreover, as the entire literature is based on these concepts and in order to understand and synthesize this literature, the interested reader must grapple with these concepts at some point. The intent of the following sections is to work through the formulation of single-factor and multifactor affine models of the term structure in continuous time. Much, if not all, of the subsequent development can be found in a variety of sources. Despite this fact, this material is almost invariably presented in abridged form, while leaving often quite

[6]For completeness, a brief discussion of single-factor discrete-time affine models is provided in Appendix C.

complex derivations to the reader. The following sections attempt to summarize and synthesize this material in sufficient detail that relatively few technical points will be left for the reader to ponder.

2.1 Some key interest rate relationships

The first step in dealing with these models involves establishing some key relationships between interest rates in a continuous-time setting, as they are used frequently in the following analysis. The discussion will be brief and deal only with the most fundamental concepts. For a more detailed discussion of these basic interest rate issues, see Björk (1998) or Richard (1978).

The fundamental building block in fixed-income analysis is a security termed a *risk-free pure discount bond*.[7] A pure discount bond is a contract that pays one unit of currency at its maturity date. It is termed risk-free because there is no risk of default. That is, the payment will be made with certainty. We denote the value of a risk-free pure discount bond as the function $P(t,T)$. The first argument, t, refers to the current time, while the second argument, T, represents the pure discount bond's maturity date; clearly, therefore, $t < T$. Given the contractual nature of the pure discount bond, we know that $P(T,T) = 1$. Restated in words, a pure discount bond has a value of \$1 at maturity.

The pure discount bond is important because, given the pure discount bond price for any given maturity, we can easily determine the associated *spot rate of interest* for that date. Indeed, the spot rate, which we denote $z(t,T)$, is the continuously compounded rate of return that generates the observed price of the pure discount bond. We solve for $z(t,T)$ in the following manner,

$$\begin{aligned} P(t,T)e^{(T-t)z(t,T)} &= 1, \\ \ln\left(P(t,T)e^{(T-t)z(t,T)}\right) &= 0, \\ \ln\left(e^{(T-t)z(t,T)}\right) &= -\ln P(t,T), \\ z(t,T) &= -\frac{\ln P(t,T)}{T-t}. \end{aligned} \tag{1}$$

Observe that if the value of the bond price function exceeds unity (i.e., $P(t,T) > 1$), it implies a negative value for the spot rate of interest. This is an undesirable trait in a nominal interest rate.

An important and useful concept in the modelling of interest rates is the *instantaneous interest rate*, which we will denote $r(t)$. A heuristic way to consider this concept is as the rate of interest demanded over an extremely short period of time. Practically, therefore, this is analogous to the overnight interest rate. In reality, of course, the instantaneous rate of interest does not exist; it is a theoretical construct used to

[7] The pure discount bond is also often called the zero-coupon bond.

facilitate the modelling process. Formally, we calculate it by evaluating the following limit:

$$\begin{aligned} r(t) = \lim_{T \to t} z(t,T) &= \lim_{T \to t} -\frac{\ln P(t,T)}{T-t}, \\ &= \lim_{T \to t} \underbrace{\left(-\frac{1}{P(t,T)} \frac{\partial \ln P(t,T)}{\partial T} \right)}_{\text{By L'Hopital}}, \\ &= -\frac{\partial \ln P(t,t)}{\partial t}. \end{aligned} \tag{2}$$

The instantaneous interest rate is the concept that takes an inherently discrete-time object such as the interest rate and gives it time-continuity. This permits us to use the mathematics of continuous-parameter stochastic processes in our modelling of the term structure of interest rates. We will see in the subsequent section that these mathematical tools are particularly useful in the derivation of the essential structure of the affine class.

2.2 The basic structure

The starting point for the development of the affine class is the postulation of a stochastic process for the state variables, or factors, that drive the dynamics of the term structure. These factors are the underlying source of uncertainty in the model of the term structure. The state variables may themselves be identified with various nodes on the term structure, or they may be considered unobservable variables, such as the mean of the instantaneous interest rate. They may also be linked to macroeconomic variables, such as inflation. In a single-factor term-structure setting, the factor is invariably identified with the instantaneous rate of interest. It will be useful in the development that follows, however, to consider the state variables to be abstract quantities that influence the variable of greatest importance; that is, the instantaneous rate of interest. We generally give the dynamics of the instantaneous interest rate the following form:

$$dr(t) = \underbrace{A_0 dt}_{\text{Drift}} + \underbrace{A_1 dW(t)}_{\text{Diffusion}}. \tag{3}$$

This interpretation of this *stochastic differential equation* is, at least conceptually, quite straightforward. Loosely speaking, it says that the differential change in the instantaneous rate ($dr(t)$) is composed of a *drift* or trend term, which is non-random, and a *diffusion* or variance term, which is random and includes a differential increment of a Brownian motion.[8] Thus, our instantaneous interest rate process bounces along

[8]Strictly speaking, of course, this is not true, as a stochastic differential does not actually exist. More formally, we would write this expression in integral form:

$$r(t) - r(0) = \int_0^t A_0 ds + \int_0^t A_1 dW(s),$$

where $W(s)$ is a standard, scalar Wiener process defined on $(\Omega, \mathcal{F}, \mathrm{P})$.

over time according to some kind of general trend. When talking about this general trend in our state variable, we have to be somewhat cautious. Interest rates are generally defined within an upper and lower bound. For this reason, our drift term includes a long-term mean parameter, defined as θ, and a mean-reversion parameter denoted κ. When the level of our instantaneous interest rate deviates from its long-term mean, θ, it will *revert* back to this mean at a speed governed by the parameter κ.[9] Our process is hampered in its ability to get back to its mean level by the diffusion term, which essentially shocks our process at each step in time. In this working paper, we look at two alternative specifications for the diffusion coefficient. One of them creates conditions that preclude negative interest rates, while the other permits negative interest rates with (depending, of course, on the parameter selection) a relatively low probability.[10] Clearly, this is an undesirable trait in any model of nominal interest rates. One might, therefore, reasonably ask why we would consider a model that permits, even with low probability, a negative interest rate outcome. The answer is one of pragmatism: the less-realistic model is substantially simpler to work with and eases some of the difficulties of parameter estimation.

It may not appear immediately obvious that we can describe the dynamics of the entire term structure of interest rates from the instantaneous rate of interest. This is accomplished using a no-arbitrage argument that is similar in spirit to the derivation of the Black-Scholes equation. In particular, we must develop and solve a partial differential equation. This is accomplished using an application of an important result from the stochastic calculus called Itô's theorem. The solution to this partial differential equation represents the zero-coupon bond price function. This function is an important link between the state variables and the term structure of interest rates. The next, likely quite natural, question is *"what is the nature of this relationship?"* The actual functions themselves are relatively simple, if somewhat lengthy, combinations of the parameters appearing in the equation describing our state variable(s). As we work through the derivation, however, a new class of parameters is introduced that describes what is called the *market price of risk*. Operationally, in our model, these parameters describe the typically upward-sloping shape of the term structure of interest rates.[11] In a general sense, these functions can take many forms and, indeed, the solution to the partial differential equation we derive may not exist. It can be shown, however, that if we place certain conditions on the coefficients A_0 and A_1, solutions to the partial differential equation do, in fact, exist and they are unique.[12] Specifically, the coefficients A_0 and A_1 must have what is termed an *affine* form. Mathematically speaking, affine means linear plus a constant, and this is the rather arcane origin of the eponymous class of

[9] Or, in symbols, we let $A_0 = \kappa(\theta - r(t))$, where $\kappa, \theta \in \mathbb{R}$.

[10] More specifically, we let $A_1 = \sigma r(t)^\gamma$, where $\gamma = 0$, which permits negative interest rates or $\gamma = \frac{1}{2}$, which precludes them. Also, we restrict $\sigma \in \mathbb{R}^+$.

[11] These parameters do not preclude the term structure from flattening or even inverting.

[12] The definition of this class of models, as well as the associated uniqueness and existence theorems, were introduced by Duffie and Kan (1996).

affine term-structure models.[13] The larger point is that for two quite reasonable choices, for these A_0 and A_1 coefficients, we are able to find closed-form solutions for the corresponding partial differential equation.[14] Moreover, as one introduces arbitrarily many state variables into the model, unique closed-form solutions for the corresponding partial differential equations continue to exist. In this way, we can incorporate additional state variables, thereby increasing the range of dynamics captured by our term-structure model.

The affine term-structure model is, in some sense, a recipe for calculating the zero-coupon rate of a given term to maturity, knowing only the value of the instantaneous rate of interest. In higher dimensions, the instantaneous rate of interest is merely comprised of a linear combination of the multiple underlying state variables. This provides a straightforward algorithm for simulation of the term structure. As the underlying state variables have explicitly specified stochastic processes, we can use them to simulate possible future sample paths. At each discrete future point in time, we may use our derived formulae to construct the entire term structure of interest rates. This term structure then acts as an input for our desired application. Now that we have provided an overview of the conceptual elements involved in these models, in sections 2.3, and 2.4 we consider the mathematical details.

2.3 The single-factor model

We begin with single-factor development to highlight the main concepts and then generalize to higher dimensions.[15] First, we postulate the following general process for the short-term interest rate. We further assume that the interest rate process is Markovian and that its dynamics are described by the following first-order stochastic differential equation:[16]

$$dr(t) = f(r,t)dt + \rho(r,t)dW(t), \tag{4}$$

where $W(t)$ is a standard scalar Wiener process defined on the filtered probability space, $(\Omega, \mathcal{F}, \mathbb{P})$. In other words, we are specifying the infinitesimal dynamics of the instantaneous interest rate under the empirical probability measure, $\mathbb{P}$. In equation (4), f represents the drift coefficient, while ρ is the diffusion term. We then further assume that a market exists for bonds of every possible maturity and that these markets are free of arbitrage. We also claim that each bond price, $P(t,T)$, is a function of the interest rate. That is, the bond price function has the following form:

$$P(t,T) = P(t,r,T). \tag{5}$$

[13] More formally, a function, $F : \mathbb{R}^n \to \mathbb{R}$, is called *affine* if there exist $a \in \mathbb{R}$ and $b \in \mathbb{R}^n$ such that $F(x) = a + b^T x$ for all $x \in \mathbb{R}^n$.

[14] Given that, typically, we will be using these models as a subcomponent in a large, computationally expensive simulation, the existence of closed-form solutions is extremely valuable, because they can be evaluated quickly.

[15] The following development comes from Vasicek (1977), who first solved the general single-factor model, and Björk (1998).

[16] More formally, the stochastic processes we will be considering are termed diffusion processes. Diffusion processes arise as strong solutions to a stochastic differential equation.

Moreover, assume that $P(t,r,T)$ is a $C^{1,2}$ function of its arguments.[17] For notational brevity, however, we will suppress the r argument from our bond price function; that is, we will continue to denote the bond price function as $P(t,T)$. An application of Itô's theorem yields the following equation:[18]

$$\begin{aligned} dP(t,T,r(t)) &= P_t dt + P_r dr(t) + \frac{1}{2} P_{rr} d\langle r \rangle(t), \\ &= P_t dt + P_r \left(\underbrace{f dt + \rho dW(t)}_{\text{equation (4)}} \right) + \frac{1}{2} P_{rr} \rho^2 dt, \\ &= \left(P_t + f P_r + \frac{\rho^2}{2} P_{rr} \right) dt + \rho P_r dW(t), \end{aligned} \tag{6}$$

where

$$\begin{aligned} P_t &= \frac{\partial P}{\partial t}, \\ P_r &= \frac{\partial P}{\partial r}, \\ P_{rr} &= \frac{\partial^2 P}{\partial r^2}. \end{aligned}$$

The key insight made by Black and Scholes, when considering stock options, involves the construction of a *self-financing* portfolio comprised of a contingent claim and the underlying asset.[19] One then proceeds to select the portfolio weights to eliminate the underlying source of uncertainty. This non-random, or riskless, portfolio must earn the risk-free rate to avoid creating an arbitrage opportunity. This no-arbitrage restriction implies that the portfolio dynamics can be described by a partial differential equation. Solving this partial differential equation analytically yields the famous Black-Scholes option pricing formula. Although this approach does not translate directly to the fixed-income setting, we may use a very similar approach. The primary difference is that the underlying instantaneous interest rate, $r(t)$, is not an asset that is traded in the marketplace. As a consequence, we will need an alternative approach to eliminate the source of uncertainty. We therefore select two pure discount bonds with arbitrary maturities for a pure discount bond, s_1 and s_2, and use them to construct a self-financing portfolio, which we will denote V. Moreover, we let u_1 and u_2 represent the weights of each bond in the portfolio. Therefore, we describe the return on our portfolio in the

[17] A function of two variables is $C^{1,2}$ if it is once continuously differentiable in its first argument and twice continuously differentiable in its second argument. This is a technical requirement for the application of Itô's theorem.

[18] In the following expression, we denote the *quadratic variation* process for the instantaneous interest rate as $\langle r \rangle(t)$. The quadratic variation process is an adapted, weakly increasing, $\mathbb{R}$-valued process such that $\{r^2(t) - \langle r \rangle(t), \mathcal{F}_t\}$ is a martingale on $(\Omega, \mathcal{F}, \mathbb{P})$. For a more detailed discusson of this concept, see Karatzas and Shreve (1991).

[19] A trading strategy is called *self-financing* if it does not require exogenous cash infusions or withdrawals over its lifetime.

following manner:

$$\frac{dV(t)}{V(t)} = u_1 \underbrace{\frac{dP_1(t,s_1)}{P_1(t,s_1)}}_{\text{Bond 1}} + u_2 \underbrace{\frac{dP_2(t,s_2)}{P_2(t,s_2)}}_{\text{Bond 2}}. \tag{7}$$

This expression comes from the fact that the return on our portfolio, V, is the sum of the proportional return in each of the two underlying assets, $P(t,s_1)$ and $P(t,s_2)$. Now, using equation (6), we can make the necessary substitutions,

$$\begin{aligned}
\frac{dV}{V} &= u_1 \frac{\left(P_{1,t} + fP_{1,r} + \frac{\rho^2}{2}P_{1,rr}\right)dt + \rho P_{1,r}dW(t)}{P_1} + u_2 \frac{\left(P_{2,t} + fP_{2,r} + \frac{\rho^2}{2}P_{2,rr}\right)dt + \rho P_{2,r}dW(t)}{P_2}, \qquad (8)\\
&= u_1 \left(\frac{\left(P_{1,t} + fP_{1,r} + \frac{\rho^2}{2}P_{1,rr}\right)}{P_1}dt + \frac{\rho P_{1,r}}{P_1}dW(t) \right) + u_2 \left(\frac{\left(P_{2,t} + fP_{2,r} + \frac{\rho^2}{2}P_{2,rr}\right)}{P_2}dt + \frac{\rho P_{2,r}}{P_2}dW(t) \right),\\
&= u_1 \left(\mu_{s_1}dt + \sigma_{s_1}dW(t)\right) + u_2 \left(\mu_{s_2}dt + \sigma_{s_2}dW(t)\right),\\
&= (u_1\mu_{s_1} + u_2\mu_{s_2})dt + \underbrace{(u_1\sigma_{s_1} + u_2\sigma_{s_2})}_{\text{Want to set to zero}}dW(t),
\end{aligned}$$

where

$$\mu_{s_i} = \frac{\left(P_{i,t} + fP_{i,r} + \frac{\rho^2}{2}P_{i,rr}\right)}{P_i},$$

and

$$\sigma_{s_i} = \frac{\rho P_{i,r}}{P_i},$$

for $i = 1, 2$. In a manner that is similar in spirit to the formulation of the original Black-Scholes partial differential equation, we wish to select those weights, u_1 and u_2, that eliminate the Brownian motion. Thus, we need to find values of u_1 and u_2 such that the following two equations hold:

$$\begin{aligned}
u_1 + u_2 &= 1,\\
u_1\sigma_{s_1} + u_2\sigma_{s_2} &= 0.
\end{aligned}$$

The trivial solution to this linear system is

$$\begin{aligned}
u_1 &= \frac{-\sigma_{s_2}}{\sigma_{s_1} - \sigma_{s_2}},\\
u_2 &= \frac{\sigma_{s_1}}{\sigma_{s_1} - \sigma_{s_2}}.
\end{aligned}$$

Now, if we plug this back into our original expression, equation (8), we have the subsequent result:

$$\begin{aligned}
\frac{dV}{V} &= \left(\frac{-\sigma_{s_2}}{\sigma_{s_1} - \sigma_{s_2}}\mu_{s_1} + \frac{\sigma_{s_1}}{\sigma_{s_1} - \sigma_{s_2}}\mu_{s_2} \right)dt + \left(\frac{-\sigma_{s_2}}{\sigma_{s_1} - \sigma_{s_2}}\sigma_{s_1} + \frac{\sigma_{s_1}}{\sigma_{s_1} - \sigma_{s_2}}\sigma_{s_2} \right)dW(t), \qquad (9)\\
\frac{dV}{V} &= \left(\frac{-\sigma_{s_2}\mu_{s_1}}{\sigma_{s_1} - \sigma_{s_2}} + \frac{\sigma_{s_1}\mu_{s_2}}{\sigma_{s_1} - \sigma_{s_2}} \right)dt.
\end{aligned}$$

We now arrive at the key insight. The portfolio, V, is riskless over the interval dt. This is because we have, by judicious selection of our portfolio weights, eliminated the source of uncertainty. Thus, over the differential period of time, dt, it *must* earn the risk-free rate to avoid creating an arbitrage opportunity. For example, if the portfolio earned more than the risk-free rate, a market participant could sell the risk-free asset short and take a simultaneous long position in V and thereby earn arbitrage profits. Thus, the rate of return on this portfolio $\frac{dV}{V}$ must be equal to $r(t)dt$. Or, rather, we have the following sequence of equalities:

$$\begin{aligned}
\frac{dV}{V} &= r(t)dt, \qquad (10)\\
\left(\frac{-\sigma_{s_2}\mu_{s_1}}{\sigma_{s_1}-\sigma_{s_2}} + \frac{\sigma_{s_1}\mu_{s_2}}{\sigma_{s_1}-\sigma_{s_2}}\right)dt &= r(t)dt,\\
\frac{\sigma_{s_1}\mu_{s_2}-\sigma_{s_2}\mu_{s_1}}{\sigma_{s_1}-\sigma_{s_2}} &= r(t),\\
\sigma_{s_1}\mu_{s_2}-\sigma_{s_2}\mu_{s_1} &= r(t)\sigma_{s_1}-r(t)\sigma_{s_2},\\
\sigma_{s_1}(\mu_{s_2}-r(t)) &= \sigma_{s_2}(\mu_{s_1}-r(t)),\\
\frac{\mu_{s_2}-r(t)}{\sigma_{s_2}} &= \frac{\mu_{s_1}-r(t)}{\sigma_{s_1}}.
\end{aligned}$$

What we have derived here is called the *market price of risk* and it represents the standardized excess return, over the risk-free rate, for holding a given pure discount bond.[20] The market price arises because the prices of bonds in our economy cannot be specified by the P-dynamics of the instantaneous interest rate in a unique manner. Essentially, this is because we have a continuum of pure discount bond maturities over the interval, $[t,T]$, but an insufficient number of underlying assets to compute their respective prices. The addition of the market price of risk adds enough additional structure to uniquely price all bonds.[21] A closer examination of equation (10) reveals that, for arbitrary maturities, s_1 and s_2, the market price of risk is the same. As a consequence, we can conclude that the market price of risk is constant across all maturities. This fact is the key insight required to derive our partial differential equation of interest. To see this, let us define the market price risk as $\lambda(t)$. Substitution of our original definitions, in equation (8), and some manipulation yield the desired partial differential equation. Note that the final expression, obtained in the following manipulation,

[20] This is also called the *Sharpe ratio*.

[21] Technically speaking, we are operating in an incomplete market. The market price of risk process is, in actuality, the *Radon-Nikodym*-derivative that allows us to specify an equivalent martingale measure, Q, from the set of equivalent measures, $\mathbf{Q}$. More specifically, it is exactly the quantity used to change the drift to our Brownian motion, $W(t)$, to create a new Brownian motion $\tilde{W}(t)$ such that all pure discount bonds return the instantaneous interest rate.

is entirely deterministic:

$$\begin{aligned}
\lambda(t) &= \frac{\mu_t - r(t)}{\sigma_t}, \qquad (11)\\
\mu_t - r(t) &= \lambda(t)\sigma_t,\\
\frac{\left(P_t + fP_r + \frac{\rho^2}{2}P_{rr}\right)}{P} - r(t) &= \lambda(t)\frac{\rho P_r}{P},\\
P_t + fP_r + \frac{\rho^2}{2}P_{rr} - r(t)P &= \lambda(t)\rho P_r,\\
P_t + (f - \rho\lambda(t))P_r + \frac{\rho^2}{2}P_{rr} - r(t)P &= 0.
\end{aligned}$$

Thus, we have derived the partial differential equation that describes the dynamics of any interest-rate-contingent claim assuming a single-state variable model. This is a tremendously important first step. At this point, we can obtain a Feynman-Kaĉ representation of the solution.[22] We desire, however, to be more specific about the coefficients in our original stochastic process for the interest rate. In particular, we wish to consider the case when the term structure $\{P(t,T), t \in [0,T]\}$ has the following form:

$$P(t,T) = e^{A(t,T) - B(t,T)r(t)}. \qquad (12)$$

When $A(t)$ and $B(t)$ are deterministic functions, the model is termed an *affine* term-structure model.[23] We often see A and B presented as functions of two arguments, t and T. Nevertheless, it is useful to think of A and B as being functions of t only while T is left as a fixed parameter. To this end, we introduce a change of variables. Denote $\tau = T - t$ and let $A(\tau) \equiv A(t,T)$ and $B(\tau) \equiv B(t,T)$. This should lead to greater notational simplicity. In addition, we suppress the time argument on the instantaneous interest rate (i.e., $r \equiv r(t)$) to avoid confusion. We thus can restate equation (12) in the following form,

$$P(\tau) = e^{A(\tau) - B(\tau)r}. \qquad (13)$$

The primary appeal of affine term-structure models is their desirable analytical properties. This is evidenced by the fact that we can now proceed to analytically solve our partial differential equation. *Why do we wish to find a solution?* Because it provides the link between our state variable and the bond prices. Thus, given this solution, we can construct a term structure of interest rates based only on knowledge of the current value of our state variable. To see specifically how this is done, let us reconsider equation (11), which is the general partial differential equation for all one-factor models that we just derived:

$$P_t + \mu P_r + \frac{\sigma^2}{2}P_{rr} - rP = 0, \qquad (14)$$

[22] The solution of certain classes of partial differential equations can be represented in terms of an expectation taken with respect to a Brownian motion. This type of solution is, broadly speaking, termed a Feynman-Kaĉ representation of the solution.

[23] Recall that a function, $F : \mathbb{R}^n \to \mathbb{R}$, is called *affine* if there exist $a \in \mathbb{R}$ and $b \in \mathbb{R}^n$ such that $F(x) = a + b^T x$ for all $x \in \mathbb{R}^n$.

where

$$\mu \triangleq f(r,t) - \rho\lambda(t).$$

and

$$\sigma \triangleq \rho(r,t).$$

We may now evaluate the various partial derivatives of $P(t,T)$ and substitute them into our partial differential equation. Here are the partial derivatives:

$$P_t = (-A'(\tau) + B'(\tau)r)e^{A(\tau)-B(\tau)r} = (-A'(\tau) + B'(\tau)r)P(\tau), \tag{15}$$

$$P_r = -B(\tau)e^{A(\tau)-B(\tau)r} = -B(\tau)P(\tau), \tag{16}$$

$$P_{rr} = B^2(\tau)e^{A(\tau)-B(\tau)r} = B^2(\tau)P(\tau), \tag{17}$$

where

$$A'(\tau) = \frac{\partial A(\tau)}{\partial \tau}\underbrace{\frac{\partial \tau}{\partial t}}_{=-1} = -\frac{\partial A(\tau)}{\partial t},$$

$$B'(\tau) = \frac{\partial B(\tau)}{\partial \tau}\frac{\partial \tau}{\partial t} = -\frac{\partial B(\tau)}{\partial t}.$$

We can plug these quantities into our partial differential equation, shown in equation (14), to obtain the following rather daunting expression:

$$\underbrace{(-A'(\tau) + B'(\tau)r)P(\tau)}_{\text{equation (15)}} - \mu\underbrace{B(\tau)P(\tau)}_{\text{equation (16)}} + \frac{\sigma^2}{2}\underbrace{B^2(\tau)P(\tau)}_{\text{equation (17)}} - rP(\tau) = 0, \tag{18}$$

$$(-A'(\tau) + B'(\tau)r(t)) - \mu B(\tau) + \frac{\sigma^2}{2}B^2(\tau) - r = 0,$$

$$-A'(\tau) - (1 - B'(\tau))r - \mu B(\tau) + \frac{\sigma^2}{2}B^2(\tau) = 0.$$

At this point, it is necessary to specify μ and σ such that $A(\tau)$ and $B(\tau)$ exist to solve this expression. It turns out, and this is the advantage of the class of affine term-structure models, that if μ and σ are themselves affine, then we can actually solve this partial differential equation explicitly. That is, there exists a unique solution.[24] Thus, we assume that our functions have the following general form:

$$\mu \triangleq \alpha_0 r + \alpha_1, \tag{19}$$

$$\sigma \triangleq \sqrt{\beta_0 r + \beta_1}.$$

[24] For more details on the actual existence and uniqueness theorems and corresponding proofs, see Duffie and Kan (1996, page 386).

Equations (13) and (19) represent, therefore, the key restrictions that define the class of affine term-structure models.[25] Let us substitute these general values into the last expression that we developed in equation (18) and simplify:

$$-A'(\tau) - (1 - B'(\tau))\, r - (\alpha_0 r + \alpha_1)B(\tau) + \frac{1}{2}(\beta_0 r + \beta_1)B^2(\tau) = 0, \tag{20}$$

$$-A'(\tau) - \alpha_1 B(\tau) + \frac{1}{2}\beta_1 B^2(\tau) - \left(1 - B'(\tau) + \alpha_0 B(\tau) - \frac{1}{2}\beta_0 B^2(\tau)\right) r = 0.$$

At first glance, it may not appear that we can do much with this expression. Observe, however, that because equation (20) holds for all r and τ, we can conclude that the two terms on the left-hand side must vanish. This reduces the problem to one of solving two, much simpler, ordinary differential equations, which are summarized as

$$-B'(\tau) + \alpha_0 B(\tau) - \frac{1}{2}\beta_0 B^2(\tau) = -1, \tag{21}$$

$$-A'(\tau) - \alpha_1 B(\tau) + \frac{1}{2}\beta_1 B^2(\tau) = 0.$$

We also have a boundary condition for the bond price. As we are considering a pure discount bond, it must be the case that the terminal value, occurring at time T, is unity. Therefore,

$$P(T,T) = e^{A(0) - B(0)r} = 1,$$

which in turn implies that $A(0) = B(0) = 0$. Given our two partial differential equations and their boundary conditions, we can actually solve for $A(\tau)$ and $B(\tau)$. Observe that the first equation,

$$-B'(\tau) + \alpha_0 B(\tau) - \frac{1}{2}\beta_0 B^2(\tau) = -1,$$

where

$$B(0) = 0,$$

depends only on $B(\tau)$ and is independent of $A(\tau)$. This means that we can solve this equation for $B(\tau)$ and substitute it into the second equation and solve for $A(\tau)$ through straightforward integration.

To emphasize this important concept and see exactly how this works, let us consider the Ornstein-Uhlenbeck process used in the Vasicek model. In this case, we have explicit expressions for the drift and diffusion terms. That is, we replace the general form described in equation (19) with the following, more-specific quantities:

$$\mu \triangleq \kappa(\bar{\theta} - r), \tag{22}$$

[25] We have expressed $\alpha_0, \alpha_1, \beta_0$, and β_1 as scalar values in this discussion. This is primarily for notational convenience. In actuality, the basic result still holds if these quantities are generalized to be arbitrary deterministic functions of t.

where

$$\bar{\theta} = \theta - \frac{\sigma\lambda}{\kappa}$$

and

$$\sigma \triangleq \sigma.$$

This is equivalent to defining our diffusion process in equation (4) as

$$dr(t) = \underbrace{\kappa(\bar{\theta} - r)}_{f(r,t)}\, dt + \underbrace{\sigma}_{\rho(r,t)}\, d\tilde{W}(t)$$

where $\tilde{W}(t)$ is a standard scalar Wiener process on the filter probability space, $\{\Omega, \mathcal{F}, \mathbb{Q}\}$.[26]

We can now, as before, substitute the expressions in equation (22) into equation (18) and simplify:

$$-A'(\tau) - (1 - B'(\tau))r - \kappa(\bar{\theta} - r)B(\tau) + \frac{\sigma^2}{2}B^2(\tau) = 0,$$

$$-A'(\tau) - (1 - B'(\tau))r - \kappa\bar{\theta}B(\tau) + \kappa r B(\tau) + \frac{\sigma^2}{2}B^2(\tau) = 0,$$

$$-A'(\tau) - \kappa\bar{\theta}B(\tau) + \frac{\sigma^2}{2}B^2(\tau) - (1 - B'(\tau) - \kappa B(\tau))\, r = 0.$$

As before, this reduces to two ordinary differential equations, which hold for all $t \in [0, T)$, with the following boundary conditions:

$$\begin{cases} B'(\tau) + \kappa B(\tau) = 1 \\ B(0) = 0, \end{cases} \tag{23}$$

$$\begin{cases} -A'(\tau) - \kappa\bar{\theta}B(\tau) + \frac{\sigma^2}{2}B^2(\tau) = 0 \\ A(0) = 0. \end{cases} \tag{24}$$

The first thing to observe is that equation (23) is a non-homogeneous, linear ordinary differential equation with the following solution:

$$B(\tau) = \frac{1}{\kappa}\left(1 - e^{-\kappa\tau}\right). \tag{25}$$

This is easily verified. We first observe that $B'(\tau) = e^{-\kappa\tau}$ and then substitute this into our postulated solution in equation (25):

$$B'(\tau) + \kappa B(\tau) = e^{-\kappa\tau} + \kappa\left(\frac{1}{\kappa}\left(1 - e^{-\kappa\tau}\right)\right) = e^{-\kappa\tau} + 1 - e^{-\kappa\tau} = 1, \tag{26}$$

[26] We have, in fact, through our selection of λ, chosen an equivalent martingale measure, $\mathbb{Q} \in \mathcal{Q}$, such that bond prices will be uniquely determined.

which yields the desired result.

The final task, therefore, is to find $A(\tau)$. Let us begin by integrating equation (24) over the range from t to T:

$$\int_t^T \left(-A'(T-s) - \kappa\bar{\theta}B(T-s) + \frac{\sigma^2}{2}B^2(T-s)\right) ds = 0.$$

At this point, we see that we will have an expression for $A(\tau)$ in terms of $B(\tau)$. All that remains, therefore, is to substitute the known function $B(\tau)$ and solve this integral. The following involves some tedious algebra, but our hard work does lead to the desired expression for $A(\tau)$:

$$A(T-t) - \underbrace{A(0)}_{=0} = \frac{\sigma^2}{2}\int_t^T B^2(T-s)ds - \kappa\bar{\theta}\int_t^T B(T-s)ds \tag{27}$$

$$\begin{aligned}
A(\tau) &= \frac{\sigma^2}{2}\int_t^T \left(\frac{1}{\kappa}\left(1-e^{-\kappa(T-s)}\right)\right)^2 ds - \kappa\bar{\theta}\int_t^T \left(\frac{1}{\kappa}\left(1-e^{-\kappa(T-s)}\right)\right) ds, \\
&= \frac{\sigma^2}{2\kappa^2}\int_t^T \left(1-2e^{-\kappa(T-s)}+e^{-2\kappa(T-s)}\right) ds - \bar{\theta}\int_t^T \left(1-e^{-\kappa(T-s)}\right) ds, \\
&= \frac{\sigma^2}{2\kappa^2}(T-t) - \frac{\sigma^2}{\kappa^2}\left(\underbrace{\frac{1-e^{-\kappa(T-t)}}{\kappa}}_{\text{equation (25)}}\right) + \frac{\sigma^2}{2\kappa^2}\left(\frac{1-e^{-2\kappa(T-t)}}{2\kappa}\right) - \bar{\theta}(T-t) + \bar{\theta}\left(\underbrace{\frac{1-e^{-\kappa(T-t)}}{\kappa}}_{\text{equation (25)}}\right), \\
&= \frac{\sigma^2}{2\kappa^2}(T-t) + \bar{\theta}B(\tau) - \bar{\theta}(T-t) - \frac{2\sigma^2}{4\kappa^2}B(\tau) - \frac{2\sigma^2}{4\kappa^2}B(\tau) + \frac{\sigma^2}{4\kappa}\left(\frac{1-e^{-2\kappa(T-t)}}{\kappa^2}\right), \\
&= \frac{\frac{\sigma^2}{2}(T-t) + \kappa^2\bar{\theta}B(\tau) - \kappa^2\bar{\theta}(T-t) - \frac{\sigma^2}{2}B(\tau)}{\kappa^2} - \frac{\sigma^2}{4\kappa}\left(\frac{(2-2e^{\kappa(T-t)})}{\kappa^2} - \frac{(1-e^{-2\kappa(T-t)})}{\kappa^2}\right), \\
&= \frac{(B(\tau)-(T-t))(\kappa^2\bar{\theta}-\frac{\sigma^2}{2})}{\kappa^2} - \frac{\sigma^2}{4\kappa}\left(\frac{1-2e^{-\kappa(T-t)}+e^{-2\kappa(T-t)}}{\kappa^2}\right), \\
&= \frac{(B(\tau)-\tau)(\kappa^2\bar{\theta}-\frac{\sigma^2}{2})}{\kappa^2} - \frac{\sigma^2}{4\kappa}\left(\underbrace{\frac{1}{\kappa}(1-e^{-\kappa\tau})}_{\text{equation (25)}}\right)^2, \\
&= \frac{(B(\tau)-\tau)(\kappa^2\left(\theta-\frac{\sigma\lambda}{\kappa}\right)-\frac{\sigma^2}{2})}{\kappa^2} - \frac{\sigma^2B^2(\tau)}{4\kappa}.
\end{aligned}$$

This permits us to conclude that in the Vasicek model, the bond price function has the following analytical form:

$$P(\tau, r) = e^{A(\tau)-B(\tau)r},$$

where, as we have just shown,

$$\begin{aligned}
B(\tau) &= \frac{1}{\kappa}(1-e^{-\kappa\tau}), \\
A(\tau) &= \frac{\gamma(B(\tau)-\tau)}{\kappa^2} - \frac{\sigma^2B^2(\tau)}{4\kappa},
\end{aligned} \tag{28}$$

where

$$\gamma = \kappa^2\left(\theta - \frac{\sigma\lambda}{\kappa}\right) - \frac{\sigma^2}{2}.$$

This analytic relationship permits us to use the instantaneous short-term rate of interest to characterize the entire term structure of interest rates at any given instant in time, given a parameter set that is consistent with the economy over a given period of time.

The CIR model can be solved in a similar, if somewhat less straightforward, fashion. It, correspondingly, has the underlying solution

$$\begin{aligned} P(\tau, r(t)) &= e^{A(\tau) - B(\tau)r}, \\ B(\tau) &= \frac{2(e^{\gamma\tau} - 1)}{(\gamma + \kappa + \lambda)(e^{\gamma\tau} - 1) + 2\gamma}, \\ A(\tau) &= \ln\left(\frac{2\gamma e^{\frac{(\gamma+\kappa+\lambda)\tau}{2}}}{(\gamma + \kappa + \lambda)(e^{\gamma\tau} - 1) + 2\gamma}\right)^{\frac{2\kappa\theta}{\sigma^2}}, \end{aligned} \tag{29}$$

where

$$\gamma = \sqrt{(\kappa + \lambda)^2 + 2\sigma^2}.$$

The details of this derivation are given in Appendix A because the resulting ordinary differential equations are more difficult to solve and working through them here adds little to the exposition. Nevertheless, the actual details are interesting and we know of no other source that actually trudges through the specifics of the derivation.

2.4 The multifactor model

Now that we have derived the general single-factor affine model, we can generalize it to higher dimensions. The structure of this derivation is based, in part, on Chaplin and Sharp (1993) for the general case and on Chen (1995), who provides a specialization to the two-factor case. The basic format is very similar to the one-factor case, although we need to explicitly consider the covariance structure between the underlying sources of randomness. In these models, we typically assume that the instantaneous short-term interest rate is in fact a linear combination of n correlated state variables, or factors, which we will denote $y_1, \ldots, y_n$. Thus, we have the following identity:

$$r = \sum_{i=1}^{n} y_i. \tag{30}$$

This is a common assumption among the factor models that we will consider in this paper. What differs between models is the underlying process followed by these factors. We begin with the entirely general multifactor case to find the form of the partial differential equation. We then solve this equation in the

specific case. This is because, as we will see later, there are restrictions on the solution to the multifactor CIR model. The simplest, and most illustrative, approach is to assume that each of the state variables is governed by a set of general stochastic differential equations as postulated in Vasicek (1977) and illustrated in equation (4):

$$\begin{aligned} dy_1(t) &= f_1(y_1,t)dt + \sum_{i=1}^{n} \rho(y_1,y_j,t)dW_1(t), \\ &\vdots \qquad\qquad \vdots \\ dy_n(t) &= f_n(y_n,t)dt + \sum_{i=1}^{n} \rho(y_n,y_j,t)dW_n(t). \end{aligned} \tag{31}$$

We observe that $W_i(t)$ is a standard scalar Wiener process defined on $(\Omega, \mathcal{F}, \mathbb{P})$ for $i = 1,...n$. As in section 2.3, the f_i terms represent the individual drift coefficients for each process, while the ρ_i's represent the diffusion terms. One important difference is that, as there exist multiple state variables, we allow the diffusion coefficient, $\rho(\cdot)$, to capture the covariance between the various sources of uncertainty. Simply put, our underlying state variables can be correlated in some manner. As in the univariate case, summarized in equation (5), we let the pure discount bond for maturity, T, be a function of the n underlying risk factors:

$$P(t,T) = P(t,T,y_1,...,y_n). \tag{32}$$

We can determine the differential dynamics of the zero-coupon bond price by an application of Itô's theorem:[27]

$$\begin{aligned} dP &= P_t dt + \sum_{i=1}^{n} P_{y_i} dy_i(t) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} P_{y_i y_j} d\langle y_i, y_j\rangle(t), \\ &= P_t dt + \sum_{i=1}^{n} P_{y_i} \underbrace{\left(f_i(y_i,t)dt + \sum_{j=1}^{n} \rho(y_i,y_j,t)d\bar{W}_i(t) \right)}_{\text{equation (31)}} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \rho(y_i,y_j,t)P_{y_i y_j} dt, \\ &= \left(P_t + \sum_{i=1}^{n} f_i(y_i,t)P_{y_i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \rho(y_i,y_j,t)P_{y_i y_j} \right) dt + \sum_{i=1}^{n}\sum_{j=1}^{n} \rho(y_i,y_j,t)P_{y_i} dW_i(t), \end{aligned} \tag{33}$$

where

$$P_{y_i} = \frac{\partial P}{\partial y_i},$$

and

$$P_{y_i y_j} = \frac{\partial P}{\partial y_i \partial y_j},$$

[27] In the following analysis, we also use the co-quadratic variance process ($\langle y_i, y_j\rangle(t)$) and the chain rule for stochastic integrals. For a thorough exposition of these concepts, see Karatzas and Shreve (1991).

for $i, j = 1, ..., n$. At this point, we use the same Black-Scholes type argument as in section 2.3 to derive the associated partial differential equation as we used in the univariate case. To do this, however, we need to hold a portfolio composed of a long position in a pure discount bond with maturity s_0, and n short positions in pure discount bonds with maturities $s_1, ..., s_n$, such that $s_0 \neq s_1 \neq \cdots \neq s_n$. Recall that in section 2.3 we had a single factor and required two bonds in our portfolio to eliminate the source of randomness. In this case, we have n sources of uncertainty and thus we require $n+1$ bonds to construct a riskless portfolio. As before, let us denote the return on our self-financing portfolio as V:

$$\frac{dV}{V} = \underbrace{\frac{dP(t,s_0)}{P(t,s_0)}}_{\text{Bond 1}} - \sum_{k=1}^{n} u_k \underbrace{\frac{dP(t,s_k)}{P(t,s_k)}}_{\substack{\text{Bond 2 to} \\ \text{Bond } (n+1)}} . \tag{34}$$

Now consider the return on this portfolio over a differential period of time, dt. Again, the plan is to select the weights on our portfolio $(u_1, ..., u_n)$ so that the n underlying sources of risk are eliminated. If we can accomplish this, then we have a portfolio that, to ensure the absence of arbitrage, must earn the risk-free rate. Thus, if we compute the total return on our portfolio using equation (34), we find the following expression:

$$\begin{aligned}
\frac{dV}{V}\mu_s &= \left(\mu(s_0)dt + \sum_{i=1}^{n}\sigma_i(s_0)dW_i(t)\right) - \sum_{k=1}^{n} u_k \left(\mu(s_k)dt + \sum_{i=1}^{n}\sigma_i(s_k)dW_i(t)\right), \qquad (35)\\
&= \left(\mu(s_0) - \sum_{k=1}^{n} u_k\mu(s_k)\right)dt + \sum_{i=1}^{n}\sigma_i(s_0)dW_i(t) - \sum_{k=1}^{n}\sum_{i=1}^{n} u_k\sigma_i(s_k)dW_i(t),\\
&= \left(\mu(s_0) - \sum_{k=1}^{n} u_k\mu(s_k)\right)dt + \left(\sum_{i=1}^{n}\sigma_i(s_0) - \underbrace{\sum_{i=1}^{n}\sum_{k=1}^{n} u_k\sigma_i(s_k)}_{\text{By Fubini}}\right)dW_i(t),\\
&= \left(\mu(s_0) - \sum_{k=1}^{n} u_k\mu(s_k)\right)dt + \sum_{i=1}^{n}\left(\underbrace{\sigma_i(s_0) - \sum_{k=1}^{n} u_k\sigma_i(s_k)}_{\text{Want to set to zero}}\right)dW_i(t),
\end{aligned}$$

where

$$\begin{aligned}
\mu(s_m) &= \frac{1}{P(t,m)}\left(P_{m,t} + \sum_{i=1}^{n} f_i(y_i,t)P_{m,y_i} + \frac{1}{2}\sum_{k=1}^{n}\sum_{j=1}^{n}\rho(y_k,y_j,t)P_{m,y_k y_j}\right) \qquad (36)\\
\sigma_i(s_m) &= \frac{1}{P(t,m)}\left(\sum_{k=1}^{n}\sum_{j=1}^{n}\rho(y_k,y_j,t)P_{m,y_i}\right),
\end{aligned}$$

for $i, j, k = 1, \cdots, n$ and $m = 0, ..., n$. To eliminate the randomness in this portfolio, we must select the appropriate $u_1, ..., u_k$ such that the following linear system is satisfied:

$$\sum_{k=1}^{n} u_k \sigma_i(s_k) = \sigma_i(s_0),$$

for $i = 1, ..., n$. This is more conveniently expressed in matrix notation,

$$\begin{bmatrix} \sigma_1(s_1) & \cdots & \sigma_1(s_n) \\ \vdots & \ddots & \vdots \\ \sigma_n(s_1) & \cdots & \sigma_n(s_n) \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} \sigma_1(s_0) \\ \vdots \\ \sigma_n(s_0) \end{bmatrix}. \tag{37}$$

Or, more simply, $Au = x$. The solution to this system exists if the matrix A is non-singular. Assuming that this is the case, the solution to this linear system is, trivially,

$$\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} \sigma_1(s_1) & \cdots & \sigma_1(s_n) \\ \vdots & \ddots & \vdots \\ \sigma_n(s_1) & \cdots & \sigma_n(s_n) \end{bmatrix}^{-1} \begin{bmatrix} \sigma_1(s_0) \\ \vdots \\ \sigma_n(s_0) \end{bmatrix}. \tag{38}$$

Or, $u = A^{-1}x$. If we substitute these portfolio weights into equation (35) we find that the instantaneous rate of return on our portfolio, V, is described in the following equation:

$$\frac{dV}{V} = \left(\mu(s_0) - \sum_{k=1}^{n} u_k \mu(s_k) \right) dt. \tag{39}$$

Observe that, as in the one-dimensional setting, we have cleverly selected our portfolio weights to eliminate the source of uncertainty in our portfolio over the differential period of time, dt. To avoid creating an arbitrage opportunity, therefore, this portfolio must earn the risk-free rate of return:

$$\left(\mu(s_0) - \sum_{k=1}^{n} u_k \mu(s_k) \right) dt = r dt, \tag{40}$$

$$\mu(s_0) - \sum_{k=1}^{n} u_k \mu(s_k) = r,$$

$$\sum_{k=1}^{n} u_k \mu(s_k) = \mu(s_0) - r,$$

$$\begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} \begin{bmatrix} \mu(s_1) \\ \vdots \\ \mu(s_n) \end{bmatrix} = \mu(s_0) - r.$$

At this point, there is a small algebraic trick to get to the *market price of risk* for each risk factor. In this context, the market price of risk is interpreted as the standardized excess return over the risk-free rate for

holding a bond of a specific maturity for a given state variable. We now recall that, given that a solution to our original linear system exists, we have equation (38), which is reproduced here for convenience:

$$\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} \sigma_1(s_1) & \cdots & \sigma_1(s_n) \\ \vdots & \ddots & \vdots \\ \sigma_n(s_1) & \cdots & \sigma_n(s_n) \end{bmatrix}^{-1} \begin{bmatrix} \sigma_1(s_0) \\ \vdots \\ \sigma_n(s_0) \end{bmatrix}.$$

If we transpose equation (40), substitute equation (38), and simplify, we get the following result,

$$\mu(s_0) - r = \begin{bmatrix} \mu(s_1) & \cdots & \mu(s_n) \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \tag{41}$$

$$\mu(s_0) - r = \begin{bmatrix} \mu(s_1) & \cdots & \mu(s_n) \end{bmatrix} \underbrace{\begin{bmatrix} \sigma_1(s_1) & \cdots & \sigma_1(s_n) \\ \vdots & \ddots & \vdots \\ \sigma_n(s_1) & \cdots & \sigma_n(s_n) \end{bmatrix}^{-1} \begin{bmatrix} \sigma_1(s_0) \\ \vdots \\ \sigma_n(s_0) \end{bmatrix}}_{\text{equation (38)}}.$$

At this point, an important definition is necessary. We define the market price of risk vector in the following manner:

$$\begin{bmatrix} \lambda_1 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \mu(s_1) & \cdots & \mu(s_n) \end{bmatrix} \begin{bmatrix} \sigma_1(s_1) & \cdots & \sigma_1(s_n) \\ \vdots & \ddots & \vdots \\ \sigma_n(s_1) & \cdots & \sigma_n(s_n) \end{bmatrix}^{-1}. \tag{42}$$

Observe from equations (36) and (42) that, for a given market price of risk term λ_i, the μ_i and σ_i terms are independent of the selection of s_i. As a consequence, we can conclude that each λ_i is also independent of each s_i. As the λ_i is constant across arbitrary maturities, it must therefore be constant for all maturities. Moreover, we term each λ_i as the market price of risk for the ith risk factor. In a manner analogous to the univariate case, we have shown that the market price of risk for each risk factor is fixed across all maturities. We can now express equation (41) as the following deterministic relation:

$$\mu(t) - r = \begin{bmatrix} \lambda_1 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \sigma_1(t) \\ \vdots \\ \sigma_n(t) \end{bmatrix} = \sum_{i=1}^{n} \lambda_i \sigma_i(t). \tag{43}$$

Finally, if we substitute the original expressions for $\mu(t)$ and $\sigma_1(t), ..., \sigma_n(t)$ from equation (36), we find that equation (43) is in fact our partial differential equation for arbitrary maturity t:

$$\frac{1}{P(t,T)}\left(P_t + \sum_{i=1}^{n} f_i(y_i,t)P_{y_i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho(y_i,y_j,t)P_{y_iy_j}\right) - r = \frac{1}{P(t,T)}.\left(\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}\lambda_i\rho(y_k,y_j,t)P_{y_i}\right) \tag{44}$$

$$P_t + \sum_{i=1}^{n} f_i(y_i,t)P_{y_i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho(y_i,y_j,t)P_{y_iy_j} - rP(t,T) - \left(\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}\lambda_i\rho(y_k,y_j,t)P_{y_i}\right) = 0$$

$$P_t + \sum_{i=1}^{n}\underbrace{\left(f(y_i,t) - \lambda_i\sum_{k=1}^{n}\sum_{j=1}^{n}\rho(y_k,y_j,t)\right)}_{\text{Drift term:}\mu_i}P_{y_i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\underbrace{\rho(y_i,y_j,t)}_{\text{Diffusion term:}\rho_{ij}}P_{y_iy_j} - \underbrace{\left(\sum_{i=1}^{n}y_i\right)}_{\text{equation (30)}}P(t,T) = 0.$$

This is the multidimensional analogue of equation (19) in section 2.3. Observe that the instantaneous rate, r, enters in the last term, $r = \sum_{i=1}^{n} y_i$. We now have our partial differential equation, which represents the dynamics of contingent claim on interest rates. It is, however, rather too general for our purposes. As a result, in a manner equivalent to equation (19) in section 2.3, we will specify the drift and diffusion coefficients for our model as follows:

$$\mu_i \triangleq \kappa_i(\bar{\theta}_i - y_i),$$

where

$$\bar{\theta}_i \triangleq \theta_i - \frac{\sigma_i\lambda_i}{\kappa_i},$$

and

$$\rho_{ij} \triangleq \sigma_{ij},$$

for $i, j = 1, ...n$. Thus, equation (44) is modified slightly to take the underlying form:

$$P_t + \sum_{i=1}^{n}\kappa_i(\bar{\theta}_i - y_i)P_{y_i} + \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\sigma_{ij}}{2}P_{y_iy_j} - \left(\sum_{i=1}^{n}y_i\right)P(t,T) = 0. \tag{45}$$

The next step is to solve this partial differential equation representing the n-dimensional Vasicek model, given the boundary condition $P(T,T,y_1,...y_n) = 1$, for the bond price function. This may seem a difficult task, but note that although we have added additional state variables, we still have an affine term-structure model. As a consequence, equation (45) can be solved in the same manner and has the same form as the one-dimensional case. That is, the solution has the general form

$$P(\tau,y_1,...,y_n) = e^{A(\tau)-\sum_{i=1}^{n}B_i(\tau)y_i}. \tag{46}$$

Let us work through the painful details and get to the form of the function $A(\tau)$ and $B_1(\tau), ..., B_n(\tau)$. The first step involves computation of the partial derivatives:[28]

$$P_t = \left(-A'(\tau) + \sum_{i=1} B_i'(\tau) y_i\right) e^{A(\tau) - \sum_{i=1}^n B_i(\tau) y_i} = \left(-A'(\tau) + \sum_{i=1} B_i'(\tau) y_i\right) P(\tau), \tag{47}$$

$$P_{y_i} = -B_i(\tau) e^{A(\tau) - \sum_{i=1}^n B_i(\tau) y_i} = -B_i(\tau) P(\tau), \tag{48}$$

$$P_{y_i y_j} = B_i(\tau) B_j(\tau) e^{A(\tau) - \sum_{i=1}^n B_i(\tau) y_i} = B_i(\tau) B_j(\tau) P(\tau), \tag{49}$$

for $i, j = 1, ..., n$. If we substitute these values into our partial differential equation, we have the following result:

$$P_t + \sum_{i=1}^n \kappa_i(\bar{\theta}_i - y_i) P_{y_i} + \sum_{i=1}^n \sum_{j=1}^n \frac{\sigma_{ij}}{2} P_{y_i y_j} - \sum_{i=1}^n y_i P(\tau) = 0, \tag{50}$$

$$\underbrace{\left(-A'(\tau) + \sum_{i=1}^n B_i'(\tau) y_i\right) P(\tau)}_{\text{equation (47)}} - \sum_{i=1}^n \kappa_i(\bar{\theta}_i - y_i) \underbrace{B_i(\tau) P(\tau)}_{\text{equation (48)}} + \sum_{i=1}^n \sum_{j=1}^n \frac{\sigma_{ij}}{2} \underbrace{B_i(\tau) B_j(\tau) P(\tau)}_{\text{equation (49)}} - \sum_{i=1}^n y_i P(\tau) = 0,$$

$$-A'(\tau) + \sum_{i=1}^n B_i'(\tau) y_i - \sum_{i=1}^n B_i(\tau) \kappa_i(\bar{\theta}_i - y_i) + \sum_{i=1}^n \sum_{j=1}^n \frac{\sigma_{ij}}{2} B_i(\tau) B_j(\tau) - \sum_{i=1}^n y_i = 0,$$

$$-A'(\tau) + \sum_{i=1}^n \sum_{j=1}^n \frac{\sigma_{ij}}{2} B_i(\tau) B_j(\tau) - \sum_{i=1}^n \kappa_i \bar{\theta}_i B_i(\tau) - \sum_{i=1}^n (1 - B_i'(\tau) - \kappa_i B_i(\tau)) y_i = 0.$$

Thus, we can again reduce equation (50), a high-dimensional partial differential equation, into a series of ordinary differential equations that can each be solved analytically. This is an example of the analytic convenience of the affine form:

$$B_i'(\tau) + \kappa_i B_i(\tau) = 1, \text{ for } i = 1, ..., n \tag{51}$$

$$-A'(\tau) + \sum_{i=1}^n \sum_{j=1}^n \frac{\sigma_{ij}}{2} B_i(\tau) B_j(\tau) - \sum_{i=1}^n \kappa_i \bar{\theta}_i B_i(\tau) = 0,$$

for $i, j = 1, ..., n$. The boundary conditions are defined as

$$A(0) = B_1(0) = \cdots = B_n(0) = 0.$$

As in the one-dimensional case, we may solve the first n ordinary differential equations for the following set of functions. The development in section 2.3 (equation (26)) justifies this statement,

$$B_i(\tau) = \frac{1}{\kappa_i}\left(1 - e^{-\kappa_i \tau}\right) \tag{52}$$

[28] Recall that, in section 2.3, we introduced the change of variables $\tau = T - t$. We will be using the same variable transformation in this subsection.

for $i = 1, ..., n$. We can now integrate the second expression in equation (51) to solve for the function $A(\tau)$:

$$\int_t^T A'(T-s)ds = \int_t^T \left(\sum_{i=1}^n \sum_{j=1}^n \frac{\sigma_{ij}}{2} B_i(T-s)B_j(T-s) - \sum_{i=1}^n \kappa_i \bar{\theta}_i B_i(T-s) \right) ds, \tag{53}$$

$$A(T-t) - \underbrace{A(0)}_{=0} = \int_t^T \sum_{i=1}^n \sum_{j=1}^n \frac{\sigma_{ij}}{2} B_i(T-s)B_j(T-s) - \int_t^T \sum_{i=1}^n \kappa_i \bar{\theta}_i B_i(T-s)ds,$$

$$A(\tau) = \sum_{i=1}^n \sum_{j=1}^n \underbrace{\frac{\sigma_{ij}}{2} \int_t^T B_i(T-s)B_j(T-s)ds}_{\text{Integral } \Psi} - \sum_{i=1}^n \underbrace{\kappa_i \bar{\theta}_i \int_t^T B_i(T-s)ds}_{\text{Integral } \Phi}.$$

Having split the previous expression into two separate integrals, Ψ and Φ, we can proceed to solve each in turn. We will start with integral Ψ:

$$\text{Integral } \Psi = \frac{\sigma_{ij}}{2} \int_t^T B_i(T-s)B_j(T-s)ds, \tag{54}$$

$$= \frac{\sigma_{ij}}{2} \int_t^T \frac{1}{\kappa_i} \left(1 - e^{-\kappa_i(T-s)}\right) \frac{1}{\kappa_j} \left(1 - e^{-\kappa_j(T-s)}\right) ds,$$

$$= \frac{\sigma_{ij}}{2\kappa_i\kappa_j} \int_t^T \left(1 - e^{-\kappa_i(T-s)} - e^{-\kappa_j(T-s)} + e^{-(\kappa_i+\kappa_j)(T-s)}\right) ds,$$

$$= \frac{\sigma_{ij}}{2\kappa_i\kappa_j} \left((T-t) - \underbrace{\frac{1}{\kappa_i}\left(1 - e^{-\kappa_i(T-t)}\right)}_{\text{equation (52)}} - \underbrace{\frac{1}{\kappa_j}\left(1 - e^{-\kappa_j(T-t)}\right)}_{\text{equation (52)}} + \frac{1}{\kappa_i+\kappa_j}\left(1 - e^{-(\kappa_i+\kappa_j)(T-t)}\right) \right),$$

$$= \frac{\sigma_{ij}}{2\kappa_i\kappa_j} \left(\tau - B_i(\tau) - B_j(\tau) + \frac{1}{\kappa_i+\kappa_j}\left(1 - e^{-(\kappa_i+\kappa_j)(\tau)}\right) \right).$$

We can then examine our second integral:

$$\text{Integral } \Phi = \kappa_i \bar{\theta}_i \int_t^T B_i(T-s)ds, \tag{55}$$

$$= \kappa_i \bar{\theta}_i \int_t^T \left(\frac{1}{\kappa_i}\left(1 - e^{-\kappa_i(T-s)}\right) \right) ds,$$

$$= \bar{\theta}_i \left((T-t) - \underbrace{\frac{1}{\kappa_i}\left(1 - e^{-\kappa_i(T-t)}\right)}_{\text{equation (52)}} \right) ds,$$

$$= \bar{\theta}_i (\tau - B_i(\tau)).$$

Having worked through the preceding, rather tiresome algebraic manipulations, we can solve equation (53) for the following closed-form expression representing $A(\tau)$:

$$A(\tau) = \sum_{i=1}^n \sum_{j=1}^n \frac{\sigma_{ij}}{2\kappa_i\kappa_j} \left(\tau - B_i(\tau) - B_j(\tau) + \frac{1}{\kappa_i+\kappa_j}\left(1 - e^{-(\kappa_i+\kappa_j)(\tau)}\right) \right) - \sum_{i=1}^n \bar{\theta}_i (\tau - B_i(\tau)). \tag{56}$$

With a moderate amount of rearrangement, we can also organize this expression into something that is more comparable to the one-factor case. That is, we get

$$A(\tau) = \sum_{i=1}^{n} \underbrace{\frac{\gamma_i(B_i(\tau)-\tau)}{\kappa_i^2} - \frac{\sigma_i^2 B_i^2(\tau)}{4\kappa_i}}_{\text{equation (28)}} + \sum_{\{i,j:i\neq j\}} \underbrace{\frac{\sigma_{ij}}{2\kappa_i\kappa_j}\left(\tau - B_i(\tau) - B_j(\tau) + \frac{1}{\kappa_i+\kappa_j}\left(1-e^{-(\kappa_i+\kappa_j)(\tau)}\right)\right)}_{\text{Covariance terms}},$$

where, as before,

$$\gamma_i = \kappa_i^2\left(\theta_i - \frac{\sigma_i\lambda_i}{\kappa_i}\right) - \frac{\sigma_i^2}{2}.$$

Thus, we recover the expression for the single-factor case, although, of course, we have n terms. The modification is that for each pair of state variables, we have an additional term describing their covariance. As in the univariate case, we have derived a closed-form expression that describes the entire term structure at any instant in time, given the values of the n risk-factors. To summarize, therefore, we have the following solution to the n-factor Vasicek partial differential equation:

$$P(\tau, y_1, \cdots, y_n) = e^{A(\tau) - \sum_{i=1}^{n} B_i(\tau) y_i}, \tag{57}$$

where

$$B_i(\tau) = \frac{1}{\kappa_i}\left(1-e^{-\kappa_i\tau}\right),$$

$$A(\tau) = \sum_{i=1}^{n} \frac{\gamma_i(B_i(\tau)-\tau)}{\kappa_i^2} - \frac{\sigma_i^2 B_i^2(\tau)}{4\kappa_i} + \sum_{\{i,j:i\neq j\}} \frac{\sigma_{ij}}{2\kappa_i\kappa_j}\left(\tau - B_i(\tau) - B_j(\tau) + \frac{1}{\kappa_i+\kappa_j}\left(1-e^{-(\kappa_i+\kappa_j)(\tau)}\right)\right).$$

As stated earlier, it is more difficult to deal with the multifactor CIR model in continuous time. The *Ricatti* equation that arises from the partial differential equation can be solved analytically only when the problem can be reduced to independent one-dimensional equations.[29] This implies that an analytic solution exists only when the underlying Brownian motions driving each state variable are independent. Thus, although the model ensures that interest rates cannot become negative, the desire for tractability implies that we give up the correlation between our state variables. Nevertheless, as we will be making use of the CIR model in the sequel, we will take a moment to describe its form. Thus, consider n independent state variables that follow the square-root process considered in the multifactor CIR model:

$$\begin{aligned} dy_1(t) &= \kappa_i(\theta_1 - y_1(t))dt + \sigma_1\sqrt{y_1(t)}dW_1(t), \\ &\vdots \\ dy_n(t) &= \kappa_n(\theta_n - y_n(t))dt + \sigma_n\sqrt{y_n(t)}dW_n(t), \end{aligned} \tag{58}$$

[29] As the multidimensional case is merely a sum of decoupled one-dimensional solutions, the derivation is identical to that considered in Appendix A.

where $W_1, ..., W_n$ are independent standard scalar Wiener processes defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Using the same general techniques as we applied to the Ornstein-Uhlenbeck processes in this section—with a few additional nuances described in Appendix A—the solution to the n-factor CIR partial differential equation has the form

$$P(\tau, y_1, ..., y_n) = e^{\sum_{i=1}^{n}(A_i(\tau) - B_i(\tau)y_i)}. \tag{59}$$

Again, the functions $A_i(\tau)$ and $B_i(\tau)$ are of the same form as in the one-dimensional case:

$$B_i(\tau) = \frac{2(e^{\gamma_i \tau} - 1)}{(\gamma_i + \kappa_i + \lambda_i)(e^{\gamma_i \tau} - 1) + 2\gamma_i}, \tag{60}$$

$$A_i(\tau) = \ln\left(\frac{2\gamma_i e^{\frac{(\gamma_i + \kappa_i + \lambda_i)\tau}{2}}}{(\gamma_i + \kappa_i + \lambda_i)(e^{\gamma_i \tau} - 1) + 2\gamma_i}\right)^{\frac{2\kappa_i \theta_i}{\sigma_i^2}},$$

where

$$\gamma_i = \sqrt{(\kappa_i + \lambda_i)^2 + 2\sigma_i^2}.$$

We have now completed the theoretical development of the two specific elements in the affine class under examination: the Vasicek and CIR models. Any discussion of the class of affine term-structure models, however, would be incomplete if it did not mention the existence of a set of analogous discrete-time affine term-structure models. In the discrete-time setting the derivation is more straightforward, if somewhat less elegant. It proceeds using the concept of a stochastic discount factor, or pricing kernel, which is a fundamental idea in microeconomics. A recursion relation is subsequently developed that provides the desired zero-coupon bond prices. This recursion relation is equivalent to the bond price function developed in the continuous-time setting. Appendix C describes the actual steps in the development of the discrete-time Vasicek and CIR models in a one-dimensional setting. This discussion is provided primarily for completeness, as we will restrict our attention to the continuous-time models in the following analysis of model implementation.

3 Model Implementation

In section 2, we developed the requisite theory used to represent bond prices as an affine function of the underlying state variables. In every case, this relationship was subject to a given parameter set. Unfortunately, the theory does not tell us anything about the appropriate values that must be specified for this parameter set. This is a matter of critical importance, because a model with poorly specified parameters is of little, if any, practical use. We must, therefore, turn to the econometric literature to handle this important issue. Although the literature on the parametric estimation of affine term-structure models is as large as the theoretical literature, there is as yet no consensus as to the best approach. In this overview, we consider a technique that has been introduced relatively recently to the estimation of affine term-structure models. The methodology we will be using, called the Kalman filter, exploits the theoretical affine relationship between

bond prices and the state variables to subsequently estimate the parameter set. The strength of this approach is that it allows the state variables to be unobserved quantities. Indeed, the Kalman filter is an algorithm that acts to identify the underlying, and unobserved, state variables that govern bond price dynamics.

3.1 The Kalman filter in brief

The Kalman filter technique has recently gained popularity in the affine term-structure literature as a result of the work by Duan and Simonato (1995), Lund (1997), Geyer and Pichler (1998), de Jong (1998), and Babbs and Nowman (1999). This approach is very useful in situations such as ours, where the underlying state variables are not observable. It is reasonable to inquire "how can we hope to estimate the parameter set when the underlying factors are unobservable?" The answer lies in the relationship between the bond prices and the underlying state variables. Indeed, we begin with an observed system of equations called the *measurement system*; this system represents exactly this affine relationship between market zero-coupon rates—which is a simple logarithmic transformation of the bond price function—and the state variables. We also have a second, unobserved system of equations termed the *transition system*. This system describes the dynamics of the state variables as they were formulated in the model. Together, the measurement and transition equations represent what is called the *state-space* form of the model. The Kalman filter uses this state-space formulation to recursively make inferences about the *unobserved* values of the state variables (transition system) by conditioning on the *observed* market zero-coupon rates (measurement system). As a final step, we use these recursive inferences to construct and maximize a log-likelihood function to find the optimal parameter set. This is a sketchy discussion, but the idea should become clearer in the discussion to follow. We begin a more detailed presentation of the Kalman filter in the subsequent section with the specifics of the state-space formulation.[30]

3.2 The state-space formulation

The reformulation of our model in what is called *state-space form* involves the specification of our measurement system and our transition system. This must be repeated for the one-, two-, and three-factor models. As the development in each case is identical, we will focus on the most difficult cases, the three-factor Vasicek and CIR models. We begin the development with the three-factor Vasicek model, by using a sequence of n zero-coupon rates $z_1, z_2, ..., z_n$ with terms to maturity denoted $t_{z_1}, t_{z_2}, ..., t_{z_n}$ respectively. In general, we only require one market zero-coupon rate for each factor used in the estimation. That is, if we were considering a two-factor model, we would require only two observed zero-coupon yields. By adding

[30]For a good overview of the filtering problem, see Oksendal (1995, chapter 6). The sources used in this paper for the Kalman filter algorithm include Kim and Nelson (1999), Hamilton (1994), Harvey (1990), and Judge, Griffiths, Hill, Lütkepohl, and Lee (1985, Appendix C).

market rates, however, we provide cross-sectional information about the term structure of interest rates at each observed point in time. This information is particularly helpful in specifying the market price of risk parameters ($\lambda_i, i = 1,2,3$). To construct our measurement system, we need these n zero-coupon rates and the following relationship between the zero-coupon yield and the price of a zero-coupon bond, which comes from section 2.4:[31]

$$z(t,T) = -\frac{\ln P(t,T)}{T-t} = \underbrace{\frac{-A(t,T) + \sum_{i=1}^{3} B_i(t,T) y_{i,t}}{T-t}}_{\text{See equation (57)}}. \tag{61}$$

Before we begin, it is important to describe how we will be discretizing the time dimension. First, we evenly subdivide the interval $[0,T]$ into N subinterval and let $t_i = i\frac{T}{N}$ for $i = 1,...,n$. Second, we denote each time-step as $\Delta t = t_i - t_{i-1}$. This discretization and equation (61) allow us to represent the observation (or measurement equation) as the following system:

$$\underbrace{\begin{bmatrix} z(t_i,t_{z_1}) \\ z(t_i,t_{z_2}) \\ \vdots \\ z(t_i,t_{z_n}) \end{bmatrix}}_{z_{t_i}} = \underbrace{\begin{bmatrix} \frac{-A(t_i,t_{z_1})}{t_{z_1}-t_i} \\ \frac{-A(t_i,t_{z_2})}{t_{z_2}-t_i} \\ \vdots \\ \frac{-A(t_i,t_{z_n})}{t_{z_n}-t_i} \end{bmatrix}}_{A} + \underbrace{\begin{bmatrix} \frac{B_1(t_i,t_{z_1})}{t_{z_1}-t_i} & \frac{B_2(t_i,t_{z_1})}{t_{z_1}-t_i} & \frac{B_3(t_i,t_{z_1})}{t_{z_1}-t_i} \\ \frac{B_1(t_i,t_{z_2})}{t_{z_2}-t_i} & \frac{B_2(t_i,t_{z_2})}{t_{z_2}-t_i} & \frac{B_3(t_i,t_{z_2})}{t_{z_2}-t_i} \\ \vdots & \vdots & \vdots \\ \frac{B_1(t_i,t_{z_n})}{t_{z_n}-t_i} & \frac{B_2(t_i,t_{z_n})}{t_{z_n}-t_i} & \frac{B_3(t_i,t_{z_n})}{t_{z_n}-t_i} \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} y_1(t_i) \\ y_2(t_i) \\ y_3(t_i) \end{bmatrix}}_{y_{t_i}} + \underbrace{\begin{bmatrix} \nu_1(t_i) \\ \nu_2(t_i) \\ \vdots \\ \nu_n(t_i) \end{bmatrix}}_{\nu_{t_i}}, \tag{62}$$

or

$$z_{t_i} = A + H y_{t_i} + \nu_{t_i},$$

where

$$\nu_t \sim \mathcal{N}(0,R),$$

$$R = \begin{bmatrix} r_1^2 & 0 & \cdots & 0 \\ 0 & r_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_n^2 \end{bmatrix}.$$

The transition equations are slightly more involved. The stochastic differential equations, which represent the dynamics of our state variables, are specified in continuous time. Thus we require some method for describing their evolution over discrete-time intervals. While there are a few alternatives—such as an Euler or Milstein scheme—we can actually solve the stochastic differential equations explicitly for y_t and then discretize this solution. While solving the stochastic differential equations in question is not, by any means,

[31] Recall that this relationship was first derived in equation (1) in section 2.2. Also note that in previous sections we have defined $\tau = T - t$. In this case, to be clear about the specific maturities of the underlying zero-coupon bonds, we will use the more descriptive notation, $T - t$.

beyond the scope of this discussion, it would take us somewhat far afield. The steps involved in the solution of the Ornstein-Uhlenbeck and square-root processes are, therefore, outlined in Appendix B. Relying on that appendix, we can specify our transition system as follows:

$$\underbrace{\begin{bmatrix} y_1(t_i) \\ y_2(t_i) \\ y_3(t_i) \end{bmatrix}}_{y_{t_i}} = \underbrace{\begin{bmatrix} \theta_1\left(1-e^{-\kappa_1\Delta t}\right) \\ \theta_2\left(1-e^{-\kappa_2\Delta t}\right) \\ \theta_3\left(1-e^{-\kappa_3\Delta t}\right) \end{bmatrix}}_{C} + \underbrace{\begin{bmatrix} e^{-\kappa_1\Delta t} & 0 & 0 \\ 0 & e^{-\kappa_2\Delta t} & 0 \\ 0 & 0 & e^{-\kappa_3\Delta t} \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} y_1(t_{i-1}) \\ y_2(t_{i-1}) \\ y_3(t_{i-1}) \end{bmatrix}}_{y_{t_{i-1}}} + \underbrace{\begin{bmatrix} \varepsilon_1(t_i) \\ \varepsilon_2(t_i) \\ \varepsilon_3(t_i) \end{bmatrix}}_{\varepsilon_{t_i}}, \tag{63}$$

or

$$y_{t_i} = C + F y_{t_{i-1}} + \varepsilon_{t_i},$$

where

$$\varepsilon_{t_i} \mid \mathcal{F}_{t_{i-1}} \sim \mathcal{N}(0, Q),$$

$$Q = \begin{bmatrix} \frac{\sigma_1^2}{2\kappa_1}\left(1-e^{-2\kappa_1\Delta t}\right) & 0 & 0 \\ 0 & \frac{\sigma_2^2}{2\kappa_2}\left(1-e^{-2\kappa_2\Delta t}\right) & 0 \\ 0 & 0 & \frac{\sigma_3^2}{2\kappa_3}\left(1-e^{-2\kappa_3\Delta t}\right) \end{bmatrix}.$$

Equations (62) and (63) together represent the state-space form of our three-factor Vasicek model. Observe that our matrix, Q, is diagonal. While the multifactor Vasicek model is sufficiently flexible to incorporate covariance between the state variables, we elected to force independence. The consequent reduction in the parameter space is helpful in ensuring the identification of model parameters and reducing the complexity of our numerical optimization algorithm.

The three-factor CIR model works in an exactly analagous manner by exploiting the fundamental relationship in equation (1). The differences relate to the slightly different error structure and form of the bond price function that is summarized in equation (60):

$$z(t,T) = -\frac{\ln P(t,T)}{T-t} = \underbrace{\frac{\sum_{i=1}^{3} -A_i(t,T) + B_i(t,T)y_{i,t}}{T-t}}_{\text{See equation (60)}}. \tag{64}$$

This allows us to represent the measurement system as,

$$\begin{bmatrix} z(t_i,t_{z_1}) \\ z(t_i,t_{z_2}) \\ \vdots \\ z(t_i,t_{z_n}) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{3} \frac{-A_i(t_i,t_{z_1})}{t_{z_1}-t_i} \\ \sum_{i=1}^{3} \frac{-A_i(t_i,t_{z_2})}{t_{z_2}-t_i} \\ \vdots \\ \sum_{i=1}^{3} \frac{-A_i(t_i,t_{z_n})}{t_{z_n}-t_i} \end{bmatrix} + \begin{bmatrix} \frac{B_1(t_i,t_{z_1})}{t_{z_1}-t_i} & \frac{B_2(t_i,t_{z_1})}{t_{z_1}-t_i} & \frac{B_3(t_i,t_{z_1})}{t_{z_1}-t_i} \\ \frac{B_1(t_i,t_{z_2})}{t_{z_2}-t_i} & \frac{B_2(t_i,t_{z_2})}{t_{z_2}-t_i} & \frac{B_3(t_i,t_{z_2})}{t_{z_2}-t_i} \\ \vdots & \vdots & \vdots \\ \frac{B_1(t_i,t_{z_n})}{t_{z_n}-t_i} & \frac{B_2(t_i,t_{z_n})}{t_{z_n}-t_i} & \frac{B_3(t_i,t_{z_n})}{t_{z_n}-t_i} \end{bmatrix} \begin{bmatrix} y_1(t_i) \\ y_2(t_i) \\ y_3(t_i) \end{bmatrix} + \begin{bmatrix} \nu_1(t_i) \\ \nu_2(t_i) \\ \vdots \\ \nu_n(t_i) \end{bmatrix}, \tag{65}$$

or

$$z_{t_i} = A + H y_{t_i} + \nu_{t_i},$$

where

$$\nu_{t_i} \sim \mathcal{N}(0, R),$$

$$R = \begin{bmatrix} r_1^2 & 0 & \cdots & 0 \\ 0 & r_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_n^2 \end{bmatrix}.$$

The form of the transition system for the CIR model is almost identical to that used for the Vasicek model. The difference arises because the state variables are governed by a set of square-root processes, rather than the Vasicek model's Ornstein-Uhlenbeck processes. This implies that the matrix, $Q \equiv Q_{t_i}$, is dependent on the state of the process. In particular, the conditional variance of the transition system has the following form:

$$\varepsilon_{t_i} \mid \mathcal{F}_{t_{i-1}} \sim \mathcal{N}(0, Q_{t_i}),$$

$$Q_{t_i} = \begin{bmatrix} \xi_1 & 0 & 0 \\ 0 & \xi_2 & 0 \\ 0 & 0 & \xi_3 \end{bmatrix},$$

and

$$\xi_j = \frac{\theta_j \sigma_j^2}{2\kappa_j} \left(1 - e^{-\kappa_j \Delta t}\right)^2 + \frac{\sigma_j^2}{\kappa_j} \left(e^{-\kappa_j \Delta t} - e^{-2\kappa_j \Delta t}\right) y_j(t_{i-1})$$

for $j = 1, 2, 3$.

Thus far, we have described the state-space representation of the three-factor CIR and Vasicek term-structure models. We will proceed to use this structure in the following section to outline the steps in the application of the Kalman filter to our estimation problem.

3.3 The Kalman filter in detail

Now that we have placed our models in state-space form, we can construct the Kalman filter for the three-factor model.[32] To provide some insight into the technique, we will describe the necessary steps in some detail. The Kalman filter originated in the engineering control literature to solve what is termed the *filtering problem*.[33] The idea is that one *observes* a stream of data over time that is subject to noise. This noise generally stems from measurement error arising in the devices used to measure the signal. In our context, the

[32] Note that the consecutive steps in the Kalman filter algorithm apply equally to both the Vasicek and CIR models.

[33] The Kalman filter is currently used in a wide variety of engineering systems, including navigational and guidance systems, the determination of satellite orbits, and radar tracking.

noise in zero-coupon rate observations might relate to bid-ask spreads, data-entry errors, or non-simultaneous observations. In addition, the observed, measured values generally depend on some set of other, unobserved, state variables. The obvious need—and the solution to the filtering problem—is to find a method for *filtering* out the desired true signal and the unobserved components from this unwanted noise. This is where the Kalman filter technique is used. Essentially, it is a recursive algorithm. It begins with an educated guess as to the initial values for the state variables and a measure of the certainty of this guess; in our case, we use the unconditional mean and variance of our state variables. The Kalman filter technique then proceeds to use these initial state variable values to infer the value of the measurement equation. Operationally, this is the expectation and variance of the measurement equation conditioned on the given initial value of the state system. The linearity assumption of the Kalman filter permits the computation of these conditional moments. Specifically, the assumption of linearity implies the equivalence of conditional expectation and orthogonal projection which, in turn, yields convenient mathematical expressions for the recursion. At this point, armed with our prediction for the measurement system, we actually observe its value. Using this observed value, we can then update our inferences about the current value of the transition system. These updated values are then used to predict the subsequent value of the state variables. We then repeat the process for the next time period. In this manner, we recurse through the entire data sample and construct a time series for our unobserved state variables. The following five steps outline the specific expressions used in the Kalman filter recursion. Of note, in the subsequent discussion, we define $\mathcal{F}_s$ as the filtration generation by the measurement system. Or, more formally,

$$\mathcal{F}_{t_i} = \sigma\{z_0, z_1, ..., z_i\},$$

where $t_i = i\frac{T}{N}$ on the interval $[0, T]$.

Step 1: Initializing the state vector. The first task is to find the appropriate starting values for the recursion. As stated previously, we use the unconditional mean and variance of our transition system. The unconditional mean, for both the CIR and Vasicek models, has the following form:

$$\mathbf{E}[y_1] = \mathbf{E}[y_1 \mid \mathcal{F}_0] = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \end{bmatrix}^T, \tag{66}$$

The unconditional variance for the Vasicek model is

$$\mathbf{var}[y_1] = \mathbf{var}[y_1 \mid \mathcal{F}_0] = \begin{bmatrix} \frac{\sigma_1^2}{2\kappa_1} & 0 & 0 \\ 0 & \frac{\sigma_2^2}{2\kappa_2} & 0 \\ 0 & 0 & \frac{\sigma_3^2}{2\kappa_3} \end{bmatrix}, \tag{67}$$

while, for the CIR model, it takes the following form:

$$\mathbf{var}[y_1] = \mathbf{var}[y_1 \mid \mathcal{F}_0] = \begin{bmatrix} \frac{\sigma_1^2 \theta_1}{2\kappa_1} & 0 & 0 \\ 0 & \frac{\sigma_2^2 \theta_2}{2\kappa_2} & 0 \\ 0 & 0 & \frac{\sigma_3^2 \theta_3}{2\kappa_3} \end{bmatrix}. \tag{68}$$

Step 2: Forecasting the measurement equation. The conditional forecast of the measurement equation has the following form:

$$\mathbf{E}\left[z_{t_i} \mid \mathcal{F}_{t_{i-1}}\right] = A + H\mathbf{E}[y_{t_i} \mid \mathcal{F}_{t_{i-1}}]. \tag{69}$$

The associated conditional variance is,

$$\mathbf{var}\left[z_{t_i} \mid \mathcal{F}_{t_{i-1}}\right] = H\mathbf{var}[y_{t_i} \mid \mathcal{F}_{t_{i-1}}]H^T + R \tag{70}$$

Step 3: Updating the inference about the state vector. We now observe the true value of the measurement system, z_{t_i}. This gives us a sense of the error in our conditional prediction, which we denote as

$$\zeta_{t_i} = z_{t_i} - \mathbf{E}[z_{t_i} \mid \mathcal{F}_{t_{i-1}}]. \tag{71}$$

At this point in the Kalman filter algorithm, this *prediction error* is used to update our inference about the unobserved transition system. This updating takes the form of revising our conditional expectation with the underlying expression

$$\mathbf{E}[y_{t_i} \mid \mathcal{F}_{t_i}] = \mathbf{E}[y_{t_i} \mid \mathcal{F}_{t_{i-1}}] + K_{t_i}\zeta_{t_i}, \tag{72}$$

where

$$K_{t_i} = \mathbf{var}[y_{t_i} \mid \mathcal{F}_{t_{i-1}}]H^T\mathbf{var}[z_{t_i} \mid \mathcal{F}_{t_{i-1}}]^{-1}, \tag{73}$$

is called the Kalman *gain matrix*. The gain matrix determines the weight given to the new observation (as summarized by the prediction error, ζ_{t_i}) in the updated state system forecast. We may also update our conditional variance of the state system using

$$\mathbf{var}[y_{t_i} \mid \mathcal{F}_{t_i}] = (I - K_{t_i}H)\,\mathbf{var}[y_{t_i} \mid \mathcal{F}_{t_{i-1}}]. \tag{74}$$

Step 4: Forecasting the state vector. In this step of the recursive loop, we may forecast the unknown values of our state system for the next time period conditioning on the updated values for the previous period. The conditional expectation is

$$\mathbf{E}[y_{t_{i+1}} \mid \mathcal{F}_{t_i}] = C + F\underbrace{\mathbf{E}[y_{t_i} \mid \mathcal{F}_{t_i}]}_{\text{equation (72)}}, \tag{75}$$

and the conditional variance has the following form:

$$\mathbf{var}[y_{t_{i+1}} \mid \mathcal{F}_{t_i}] = \mathbf{var}[y_{t_i} \mid \mathcal{F}_{t_{i-1}}] - F \underbrace{\mathbf{var}[y_{t_i} \mid \mathcal{F}_{t_i}]}_{\text{equation (74)}} F^T + Q, \tag{76}$$

Step 5: Constructing the likelihood function. The previous four steps must be repeated for each discrete time step in the data sample. In our analysis, we use monthly data over a period of ten years. To actually implement this algorithm to estimate the parameter set, we initialize the state vector using equation (66) and (68) and then iterate on equations (69) to (76). At each step, we generate a measurement-system prediction error (ζ_{t_i}) and a prediction error covariance matrix ($\mathbf{var}[y_{t_i} \mid \mathcal{F}_{t_{i-1}}]$). Under the assumption that our measurement-system prediction errors are Gaussian, we can construct the log-likelihood function. It will have the following usual form

$$\begin{aligned} \ell(\theta) &= \sum_{i=1}^{N} \ln\left[(2\pi)^{-\frac{2}{2}} \det\left(\mathbf{var}[y_{t_i} \mid \mathcal{F}_{t_{i-1}}]\right)^{-\frac{1}{2}} e^{-\frac{1}{2}\zeta_{t_i}^T \mathbf{var}[y_{t_i} \mid \mathcal{F}_{t_{i-1}}]^{-1}\zeta_{t_i}}\right], \\ &= -\frac{nN\ln(2\pi)}{2} - \frac{1}{2}\sum_{i=1}^{N}\left[\ln\left(\det\left(\mathbf{var}[y_{t_i} \mid \mathcal{F}_{t_{i-1}}]\right)\right) + \zeta_{t_i}^T \mathbf{var}[y_{t_i} \mid \mathcal{F}_{t_{i-1}}]^{-1}\zeta_{t_i}\right]. \end{aligned} \tag{77}$$

In other words, steps one through four are a lengthy procedure for the construction of a log-likelihood function in disguise. To find the optimal parameter set, therefore, we merely treat the preceding algorithm as our objective function and use non-linear numerical optimization techniques to find the maximum.

3.4 Simulation results

In this section, we apply the preceding theoretical discussion of our estimation technique to the problem at hand. That is, we place the one-, two-, and three-factor Vasicek and CIR models into state space form. We then simulate various term-structure outcomes using a known parameter set and proceed to estimate the model parameters. This simulation exercise is intended to indicate how effective this technique is in terms of identifying parameters. In particular, we follow the subsequent sequence of steps in the simulation exercise:

- The first step is to simulate our term-structure paths. The state variables are simulated from the discretized solution to their attendant stochastic differential equation.[34] In the estimation, we used monthly observations over a 10-year time horizon. The state variables, however, are actually simulated over a finer time discretization. In particular, we compute the value of the state variables weekly over the ten-year period, but use only the monthly observations in our data. The rationale behind this approach is to achieve a better approximation of the underlying stochastic processes that govern our

[34] Recall that the technical details are described in Appendix B.

state variables. As a final note, we assume that the zero-coupon rates are observed with a normally distributed independent error term; technically, this implies that our matrix, R, is diagonal.[35]

- Starting with an arbitary set of parameters for each term-structure sample path, we then proceed to run the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization algorithm in an effort to find the optimal parameter set. The actual optimization routine does not constrain the parameter values, but instead the variables are passed through a continuous transformation to effectively constrain their values. In particular, if we denote c as the unconstrained variable and $\bar{c}$ as the constrained variable, then the transformation is of the form

$$\bar{c} = \delta\left(\frac{e^c}{1+e^c}\right) \tag{78}$$

where $\delta \in \mathbb{R}$. The scalar value δ determines the size interval over which the constrained parameter may vary. For the mean-reversion parameters ($\kappa_i, i = 1,2,3$), δ is equal to one, while for the volatility ($\sigma_i, i = 1,2,3$) and long-term mean ($\theta_i, i = 1,2,3$) parameters, δ is set to $\frac{1}{4}$. For the market price of risk parameters ($\lambda_i, i = 1,2,3$), we let δ be equal to -1.

- The simulation of a sample path for the term structure of interest rates, followed by application of the estimation algorithm, is repeated 250 times. This may not be an entirely sufficient number of iterations, but the procedure is rather lengthy and the results are primarily intended to give a general sense of the accuracy of the approach.

The following tables summarize the results of the simulation exercise for the Vasicek and CIR models. They report the true values, the mean estimate over the 250 simulations, and the associated standard deviation of the estimates. The mean estimate value provides some sense of any bias in the estimation technique, while the corresponding standard deviation is useful in assessing the accuracy of the approach.

In reviewing Tables 1 through 3, we can make the following specific observations.

- The Kalman filter tends to work quite well for each of the three separate Vasicek models. In particular, it appears to identify quite closely the mean-reversion, long-term mean, and volatility parameters. The only problems that do arise tend to be restricted to the market price of risk parameters. While the estimates of these parameters do not appear to be biased, the standard errors are quite large in both relative and absolute terms.

- The results, while not terrible, are slightly less encouraging for the three CIR models. The most striking result is that there appears to be a sizable upward bias in the estimated value of the mean-reversion

[35] In fact, assuming that we have k zero-coupon bonds in our measurement system, we simulate the measurement errors as i.i.d. $\tau_i \sim \mathcal{N}(0, 0.001^2)$ for $i = 1, ..., k$.

Table 1: **Simulation Exercise, One-Factor Models**: This table summarizes the results of the simulation exercise for the one-factor Vasicek and CIR models. In both instances, 250 estimations were performed using 1-month, 3-month, 6-month, and 10-year bonds.

	Vasicek			CIR		
Parameters	**Actual values**	**Mean estimate**	**Standard deviation**	**Actual values**	**Mean estimate**	**Standard deviation**
κ	0.06	0.062	0.018	0.10	0.141	0.053
θ	0.05	0.048	0.025	0.05	0.041	0.013
σ	0.02	0.020	0.001	0.075	0.075	0.005
λ	-0.20	-0.204	0.079	-0.40	-0.437	0.042

Table 2: **Simulation Exercise, Two-Factor Models**: This table summarizes the results of the simulation exercise for the two-factor Vasicek and CIR models. In both instances, 250 estimations were performed using nine different bond prices ranging from 3 months to 10 years.

	Vasicek			CIR		
Parameters	**Actual values**	**Mean estimate**	**Standard deviation**	**Actual values**	**Mean estimate**	**Standard deviation**
κ_1	0.06	0.060	0.005	0.10	0.170	0.122
κ_2	0.70	0.700	0.011	0.70	0.791	0.137
θ_1	0.05	0.053	0.022	0.05	0.034	0.017
θ_2	0.01	0.010	0.007	0.03	0.035	0.013
σ_1	0.02	0.020	0.001	0.075	0.079	0.013
σ_2	0.05	0.050	0.003	0.05	0.054	0.012
λ_1	-0.20	-0.189	0.067	-0.20	-0.274	0.123
λ_2	-0.50	-0.504	0.251	-0.10	-0.170	0.138

and market price of risk parameters throughout all CIR models; moreover, the estimates for these parameters have rather substantial standard errors. Visual inspection of equation (60) reveals one possible explanation: the κ and λ parameters that occur in the $A(\tau)$ and $B(\tau)$ terms almost invariably appear jointly. This might create some difficulty in their identification.

- In general, all of the Kalman filter does a reasonable job of estimating the volatility and long-term

Table 3: **Simulation Exercise, Three-Factor Models**: This table summarizes the results of the simulation exercise for the three-factor Vasicek and CIR models. In both instances, 250 estimations were performed using 14 different bond prices ranging from one month to 30 years.

	Vasicek			CIR		
Parameters	**Actual values**	**Mean estimate**	**Standard deviation**	**Actual values**	**Mean estimate**	**Standard deviation**
κ_1	0.06	0.061	0.019	0.25	0.296	0.100
κ_2	0.30	0.294	0.061	0.45	0.514	0.158
κ_3	0.70	0.699	0.085	0.80	0.797	0.152
θ_1	0.01	0.005	0.005	0.05	0.042	0.019
θ_2	0.02	0.019	0.014	0.03	0.034	0.019
θ_3	0.04	0.071	0.021	0.01	0.014	0.014
σ_1	0.02	0.021	0.011	0.05	0.057	0.018
σ_2	0.05	0.050	0.009	0.075	0.078	0.031
σ_3	0.03	0.031	0.009	0.15	0.137	0.022
λ_1	-0.20	-0.163	0.143	-0.15	-0.193	0.078
λ_2	-0.50	-0.420	0.240	-0.10	-0.125	0.067
λ_3	-0.15	-0.263	0.267	-0.05	-0.074	0.051

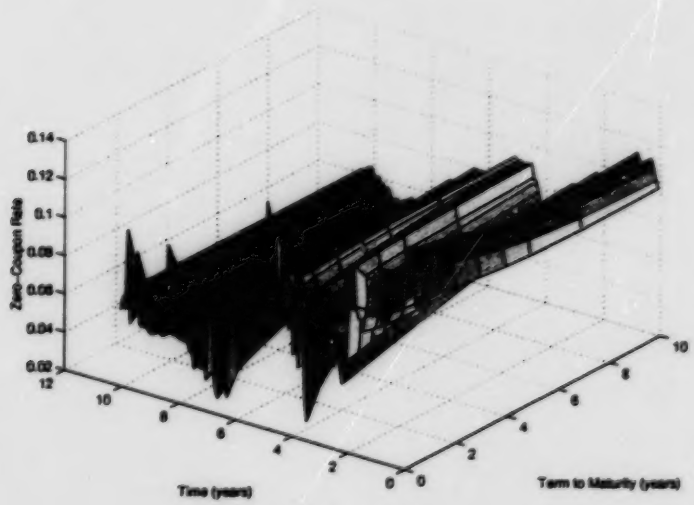
mean parameters in all the models. The one exception is the θ_1 parameter in the two-factor CIR model.

Overall, based on the results of the simulation exercise, we may cautiously conclude that the Kalman filter is a successful technique for determining the parameters in our two affine term-structure models. While we can strengthen this statement somewhat when speaking of the Vasicek model, the accuracy of our estimation technique appears to deteriorate as we increase the number of state variables. This should not be too surprising, given that we are placing an increasing burden on the Kalman filter to identify the unobserved factors.

3.5 Actual results

In this section, we apply our estimation technique to some actual Canadian zero-coupon data. While zero-coupon rates are not themselves directly observable, we use ten years of monthly data generated from the Svensson model as implemented by Bolder and Stréliski (1999). These data correspond to the fifteenth day of each month, from January 1990 to June 2000. Figure 3 plots the evolution of the actual Canadian term structure of interest rates as estimated by the Svensson model; this can be compared to the stylized illustration of the term structure summarized in Figure 2. The market zero-coupon rates incorporated into our measurement system will, therefore, include twelve observations with 6-month, 1-year, $1\frac{1}{2}$-year, 2-year, 3-year, 4-year, 5-year, 7-year, 10-year, 15-year, 20-year, and 30-year terms to maturity. As in the simulation exercise, we assume that the zero-coupon rates are observed with independent normally distributed errors. Operationally, this implies that the off-diagonal elements of the R system matrix are zero.

Figure 3: **The Actual Empirical Term Structure**: These term-structure data were produced using the Svensson model as detailed in Bolder and Stréliski (1999).



The steps followed in the optimization algorithm are very similar to those followed in the previous simulation exercise. In particular, we start with an arbitrary parameter set, use an unconstrained BFGS optimization algorithm, and perform the same continuous transformation of our parameter values. The standard errors, however, are computed from the Fisher information matrix. If we have k parameters and

we denote the vector of standard errors as ψ, then we compute each individual standard error as follows:

$$\psi_i = \sqrt{H_{ii}^{-1}} \tag{79}$$

for $i = 1, ..., k$ where H is the Hessian matrix.

Table 4 summarizes the specific results for the one-factor Vasicek and CIR models. It is interesting to note that the two models have very similar long-term mean values, but exhibit both substantially different mean reversion speeds and market prices of risk. In addition, both estimates of the market price of risk have relatively large standard errors.

Table 4: **Actual Data, One-Factor Models**: This table includes the parameter estimates—performed using actual Canadian zero-coupon data—and their associated standard errors for the one-factor Vasicek and CIR models.

	Vasicek		CIR	
Parameters	**Estimate**	**Std. error**	**Estimate**	**Std. error**
κ	0.147	0.025	0.655	0.012
θ	0.074	0.005	0.073	0.005
σ	0.029	0.004	0.136	0.008
λ	-0.154	0.072	-0.313	0.124

The results of the two- and three-factor models in Tables 5 and 6 are more difficult to interpret. The Vasicek model tends to keep the sum of the long-term mean variables close to 7 per cent, which is consistent with the one-factor long-term mean estimate; the CIR model, however, suggests a long-term mean closer to 10 per cent, which seems to be overstated. Another interesting point is that there seems to be a positive relationship between mean-reversion and volatility. That is, the larger a factor's mean-reversion coefficient, the higher its corresponding volatility parameter. This is particularly evident with the CIR model. As a final note, the standard error values seem to be substantially smaller than those observed in the previous simulation exercise. This may be cause to question their accuracy.

This paper does not, however, have an empirical focus, and these results are primarily illustrative. Our goal at the outset was to consider the practical details in estimating the parameter set for the models examined in a theoretical light in section 2. To that end, we have worked through the requisite details of the algorithm, conducted a simulation exercise to ascertain the accuracy of this technique, and then applied our method to actual data.

Table 5: **Actual Data, Two-Factor Models**: This table includes the parameter estimates—performed using actual Canadian zero-coupon data—and their associated standard errors for the two-factor Vasicek and CIR models.

	Vasicek		CIR	
Parameters	**Estimate**	**Std. error**	**Estimate**	**Std. error**
κ_1	0.043	0.058	0.024	0.036
κ_2	0.376	0.060	0.595	0.037
θ_1	0.060	0.004	0.050	0.004
θ_2	0.009	0.003	0.052	0.005
σ_1	0.015	0.008	0.111	0.008
σ_2	0.017	0.008	0.150	0.007
λ_1	-0.045	0.101	-0.148	0.029
λ_2	-0.253	0.314	-0.026	0.022

Table 6: **Actual Data, Three-Factor Models**: This table includes the parameter estimates—performed using actual Canadian zero-coupon data—and their associated standard errors for the two-factor Vasicek and CIR models.

	Vasicek		CIR	
Parameters	**Estimate**	**Std. error**	**Estimate**	**Std. error**
κ_1	0.019	0.022	0.052	0.034
κ_2	0.361	0.037	0.235	0.033
κ_3	0.958	0.031	0.862	0.028
θ_1	0.004	0.002	0.008	0.003
θ_2	0.014	0.003	0.061	0.003
θ_3	0.064	0.004	0.037	0.003
σ_1	0.015	0.004	0.088	0.005
σ_2	0.062	0.005	0.111	0.005
σ_3	0.079	0.005	0.123	0.005
λ_1	-0.063	0.038	-0.168	0.016
λ_2	-0.279	0.035	-0.084	0.011
λ_3	-0.561	0.033	-0.037	0.006

4 Conclusion

The first part of this paper focused on the theoretical development, in continuous time, of the class of affine term-structure models. This development begins with a single-factor model where the starting point is the postulation of a stochastic process for the underlying state variable. No-arbitrage restrictions permit the construction of a riskless portfolio, which in turn yields a partial differential equation representing the dynamics of a zero-coupon bond portfolio. This partial differential equation is then solved to generate the critical bond price function, which encapsulates the affine relationship between the state variable and the bond price. This key result is then generalized—using multiple state variables—into higher dimensions, providing us with the tools to model the intertemporal evolution of the term structure of interest rates. As term-structure models need not be constructed in continuous time, a brief exposition of the derivation of a corresponding discrete-time affine term-structure model is provided for completeness in Appendix C.

Having developed the necessary theory, the second part of this paper focused on model implementation. Implementation in this setting is operationalized as parameter estimation. Moreover, parameter estimation of affine term-structure models is essentially a time-series problem. As a consequence, a variety of techniques exist to deal with the parameter estimation problem. The technique used was selected, as it does not require *a priori* that the state variables be observable. The methodology—which comes from the engineering control literature and is termed the Kalman filter—represents the model in state-space form and uses a recursive approach to determine the optimal parameter set. In addition to explicitly detailing the steps involved in applying the Kalman filter to this problem, the implementation section provided two empirical exercises. We first performed a Monte Carlo experiment where, given a known parameter set, data were simulated and each technique was used to estimate the parameters. This is a useful diagnostic to determine the accuracy of the estimation technique. Moreover, the results indicate that this technique does a reasonably good job of estimating a given parameter set. In the second empirical exercise, we applied the Kalman filter algorithm to actual Canadian zero-coupon data.

The affine class of term-structure models is in many respects the introductory point for term-structure modelling. Affine models possess attractive analytic properties, they generalize relatively easily to higher dimensions, and thus they represent a reasonable first model of term-structure dynamics. This family of models nonetheless suffers from a number of drawbacks, including its inherent linearity and its inability to fit the observed current term structure. Interesting avenues for future research would address these two concerns. Also, it remains unknown whether these models adequately explain the dynamics of the Canadian term structure. More focused work in this area is clearly required. In conclusion, therefore, this paper represents a first step in advancing the Bank of Canada's research agenda in this area. Given the practical importance of generating reasonable term-structure models for debt strategy and risk-management analysis, term-structure modelling should remain a rich area for future exploration.

Appendix A: The One-Factor CIR Model

In this appendix, we look at the specific case of the stochastic process specified in the CIR model. The steps involved are very similar to those followed in the derivation of the Vasicek model. In this example, however, the solution is more involved. We nevertheless start in the same manner with explicit expressions for the drift and diffusion terms:

$$\begin{aligned} \mu &\triangleq \kappa\theta - (\kappa + \lambda)r, \\ \sigma &\triangleq \sigma\sqrt{r}. \end{aligned} \tag{80}$$

Observe that the key differences are that we have added a $\sqrt{r}$ term into the diffusion coefficient, and that the market price of risk depends on the level of the instantaneous rate. These changes make both the volatility and the associated market price of risk proportional to the level of our state variable. The benefit of the change to the drift term is to preclude negative interest rates.[36] We may again substitute into equation (18) and simplify:

$$\begin{aligned} -A'(\tau) - (1 - B'(\tau))\,r - (\kappa\theta - (\kappa + \lambda)r)B(\tau) + \frac{\sigma^2 r}{2}B^2(\tau) = 0, \\ -A'(\tau) - (1 - B'(\tau))\,r - \kappa\theta B(\tau) + (\kappa + \lambda)rB(\tau) + \frac{\sigma^2}{2}B^2(\tau) = 0, \\ -A'(\tau) - \kappa\theta B(\tau) - \left(1 - B'(\tau) - (\kappa + \lambda)B(\tau) - \frac{\sigma^2}{2}B^2(\tau)\right)r = 0. \end{aligned} \tag{81}$$

As before, this reduces to two ordinary differential equations with the following boundary conditions:

$$\begin{cases} -B'(\tau) - (\kappa + \lambda)B(\tau) - \frac{\sigma^2}{2}B^2(\tau) = -1, \\ B(0) = 0, \end{cases} \tag{82}$$

$$\begin{cases} -A'(\tau) - \kappa\theta B(\tau) = 0, \\ A(0) = 0. \end{cases} \tag{83}$$

This is where things get more difficult. The first ordinary differential equation, in equation (82), is called a *Ricatti* equation. In general, this type of equation must be solved using a numerical technique such as the *Runge-Kutta* method.[37] In this instance, fortunately, there turns out to be an explicit solution of the form

$$B(\tau) = \frac{2(e^{\gamma\tau} - 1)}{(\gamma + \kappa + \lambda)(e^{\gamma\tau} - 1) + 2\gamma}, \tag{84}$$

[36] The new drift term alone does not necessarily guarantee positive interest rates. If, however, we ensure that $2\kappa\theta > \sigma^2$, then negative interest rates are indeed precluded.

[37] This method is a so-called *predictor-corrector* method, which improves on the basic *Euler* discretization. For more information on this method see Press, Teukolsky, Vetterling, and Flannery (1992, page 710).

where

$$\gamma = \sqrt{(\kappa+\lambda)^2 + 2\sigma^2}.$$

Deriving this solution is a bit involved, but we can verify that it is indeed the solution. To do this, let us start by calculating the various components:

$$B'(\tau) = \frac{4\gamma^2 e^{\gamma\tau}}{\Phi^2}, \tag{85}$$

$$B(\tau) = \frac{2e^{\gamma\tau}-2}{\Phi}, \tag{86}$$

$$B^2(\tau) = \frac{4\left(e^{\gamma\tau}-1\right)^2}{\Phi^2}, \tag{87}$$

where, for notational simplicity,

$$\Phi = (\gamma+\kappa+\lambda)(e^{\gamma\tau}-1)+2\gamma.$$

As an additional note, if we square Φ and expand it, we have

$$\begin{aligned}
\Phi^2 &= ((\gamma+(\kappa+\lambda))(e^{\gamma\tau}-1)+2\gamma)^2, \qquad (88)\\
\Phi^2 &= (\gamma e^{\gamma\tau}+\gamma+(\kappa+\lambda)e^{\gamma\tau}-(\kappa+\lambda))^2,\\
\Phi^2 &= 4\sigma^2 e^{\gamma\tau}+2e^{2\gamma\tau}(\sigma^2+(\kappa+\lambda)\gamma+(\kappa+\lambda)^2)+2(\sigma^2-(\kappa+\lambda)\gamma+(\kappa+\lambda)^2).
\end{aligned}$$

To see whether this function solves our Ricatti equation, let us plug in the components and work through the tedious details of simplifying the expression to show that, indeed, we have a solution to our Ricatti equation:

$$\begin{aligned}
\text{LHS of equation (82)} &= -B'(\tau)-(\kappa+\lambda)B(\tau)-\frac{\sigma^2}{2}B^2(\tau) \qquad (89)\\
&= -\underbrace{\frac{4\gamma^2 e^{\gamma-\tau}}{\Phi^2}}_{\text{equation (85)}} -(\kappa+\lambda)\underbrace{\left(\frac{2e^{\gamma\tau}-2}{\Phi}\right)}_{\text{equation (86)}} -\frac{\sigma^2}{2}\underbrace{\left(\frac{4\left(e^{\gamma-\tau}-1\right)^2}{\Phi^2}\right)}_{\text{equation (87)}},\\
&= \frac{-4\sigma^2 e^{\gamma\tau}-2e^{2\gamma\tau}(\sigma^2+(\kappa+\lambda)\gamma+(\kappa+\lambda)^2)-2(\sigma^2-(\kappa+\lambda)\gamma+(\kappa+\lambda)^2)}{\Phi^2},\\
&= \underbrace{\frac{-4e^{\gamma\tau}(\sigma^2)-2e^{2\gamma\tau}(\sigma^2+(\kappa+\lambda)\gamma+(\kappa+\lambda)^2)-2(\sigma^2-(\kappa+\lambda)\gamma+(\kappa+\lambda)^2)}{4e^{\gamma\tau}(\sigma^2)+2e^{2\gamma\tau}(\sigma^2+(\kappa+\lambda)\gamma+(\kappa+\lambda)^2)+2(\sigma^2-(\kappa+\lambda)\gamma+(\kappa+\lambda)^2)}}_{\text{equation (88)}},\\
&= -\left(\frac{4e^{\gamma\tau}(\sigma^2)+2e^{2\gamma\tau}(\sigma^2+(\kappa+\lambda)\gamma+(\kappa+\lambda)^2)+2(\sigma^2-(\kappa+\lambda)\gamma+(\kappa+\lambda)^2)}{4e^{\gamma\tau}(\sigma^2)+2e^{2\gamma\tau}(\sigma^2+(\kappa+\lambda)\gamma+(\kappa+\lambda)^2)+2(\sigma^2-(\kappa+\lambda)\gamma+(\kappa+\lambda)^2)}\right)\\
&= -1 = \text{RHS of equation (82)}.
\end{aligned}$$

The next step is to solve the second equation. The first step, which is merely to make the integration easier to follow, is to consider the interval $[0,t]$ rather than the more general $[t,T]$.[38] This implies that $\tau = t$ and

[38] Note that our final result holds without loss of generality, as we may always normalize our time interval, $[t,T]$, to make $t=0$.

we may alter our limits of integration in the subsequent fashion:

$$\int_0^t A'(s)ds = -\kappa\theta \int_0^t B(s)ds, \tag{90}$$

$$A(t) - \underbrace{A(0)}_{=0} = -\kappa\theta \int_0^t \frac{2(e^{\gamma s}-1)}{(\gamma+\kappa+\lambda)(e^{\gamma s}-1)+2\gamma}ds,$$

$$A(t) = -2\kappa\theta \int_0^t \frac{e^{\gamma s}-1}{(\gamma+\kappa+\lambda)(e^{\gamma s}-1)+2\gamma}ds.$$

At this point, we need to introduce a change of variables. If we let

$$z = e^{\gamma s},$$

then

$$ds = \frac{1}{z\gamma}dz.$$

Also, let us simplify this expression somewhat by letting,

$$\hat{A} = \gamma+\kappa+\lambda, \tag{91}$$

$$\hat{B} = 2\gamma.$$

Using this substitution, we have an integrand that is a bit easier to work with:

$$A(t) = -2\kappa\theta \int_1^{e^{\gamma t}} \left(\frac{z-1}{\hat{A}(z-1)+\hat{B}}\right)\left(\frac{1}{z\gamma}\right)dz, \tag{92}$$

$$= -\frac{2\kappa\theta}{\gamma}\int_1^{e^{\gamma t}} \frac{z-1}{(\hat{A}(z-1)+\hat{B})z}dz,$$

$$= -\frac{2\kappa\theta}{\gamma}\left[\int_1^{e^{\gamma t}} \frac{z}{(\hat{A}(z-1)+\hat{B})z}dz - \int_1^{e^{\gamma t}} \frac{1}{(\hat{A}(z-1)+\hat{B})z}dz\right],$$

$$= -\frac{2\kappa\theta}{\gamma}\left[\int_1^{e^{\gamma t}} \frac{1}{\hat{A}(z-1)+\hat{B}}dz - \int_1^{e^{\gamma t}} \left(\frac{M}{\hat{A}(z-1)+\hat{B}}\right)+\left(\frac{N}{z}\right)dz\right],$$

where

$$M = \frac{\hat{A}}{\hat{A}-\hat{B}},$$

$$N = \frac{1}{\hat{B}-\hat{A}}.$$

We plug these back into equation (92) and simplify further:

$$A(t) = \frac{2\kappa\theta}{\gamma}\left[\int_1^{e^{\gamma t}} \frac{1}{\hat{A}(z-1)+\hat{B}}dz - \int_1^{e^{\gamma t}} \left(\frac{\frac{\hat{A}}{\hat{A}-\hat{B}}}{\hat{A}(z-1)+\hat{B}}\right)+\left(\frac{\frac{1}{\hat{B}-\hat{A}}}{z}\right)dz\right], \tag{93}$$

$$= \frac{2\kappa\theta}{\gamma}\left[\int_1^{e^{\gamma t}} \frac{1}{(\hat{A}(z-1)+\hat{B})}dz - \left(\frac{\hat{A}}{\hat{A}-\hat{B}}\right)\int_1^{e^{\gamma t}} \frac{1}{\hat{A}(z-1)+\hat{B}}dz - \left(\frac{1}{\hat{B}-\hat{A}}\right)\int_1^{e^{\gamma t}} \frac{1}{z}dz\right].$$

We now introduce a second change of variables. Let

$$u = \hat{A}(z-1) + \hat{B},$$

and observe that

$$dz = \frac{1}{\hat{A}} du.$$

Substituting this change into our lengthening expression for $A(t)$ in equation (93), we obtain the following, which we can simplify substantially:

$$\begin{aligned}
A(t) &= -\frac{2\kappa\theta}{\gamma}\left[\left(\frac{1}{\hat{A}}\right)\int_{\hat{B}}^{\hat{A}(e^{\gamma t}-1)+\hat{B}} \frac{1}{u}du - \left(\frac{\hat{A}}{(\hat{A}-\hat{B})\hat{A}}\right)\int_{\hat{B}}^{\hat{A}(e^{\gamma t}-1)+\hat{B}} \frac{1}{u}du - \left(\frac{1}{\hat{B}-\hat{A}}\right)\int_{1}^{e^{\gamma t}} \frac{1}{z}dz\right], \quad (94)\\
&= -\frac{2\kappa\theta}{\gamma}\left[\frac{1}{\hat{A}}\left[\ln u\right]_{\hat{B}}^{\hat{A}(e^{\gamma t}-1)+\hat{B}} - \frac{1}{\hat{A}-\hat{B}}\left[\ln u\right]_{B}^{A(e^{\gamma t}-1)+B} - \frac{1}{\hat{B}-\hat{A}}\left[\ln z\right]_{\hat{B}}^{\hat{A}(e^{\gamma t}-1)+\hat{B}}\right],\\
&= -\frac{2\kappa\theta}{\gamma}\left[\frac{1}{\hat{A}}\left[\ln(\hat{A}(z-1)+\hat{B})\right]_{B}^{A(e^{\gamma t}-1)+B} - \frac{1}{\hat{A}-\hat{B}}\left[\ln(\hat{A}(z-1)+\hat{B})\right]_{B}^{A(e^{\gamma t}-1)+B} - \frac{1}{\hat{B}-\hat{A}}\left[\ln(e^{\gamma s})\right]_{0}^{t}\right],\\
&= -\frac{2\kappa\theta}{\gamma}\left[\frac{1}{\hat{A}}\left[\ln(\hat{A}(e^{\gamma s}-1)+\hat{B})\right]_{0}^{t} - \frac{1}{\hat{A}-\hat{B}}\left[\ln(\hat{A}(e^{\gamma s}-1)+\hat{B})\right]_{0}^{t} - \frac{1}{\hat{B}-\hat{A}}\ln(e^{\gamma t})\right],\\
&= -\frac{2\kappa\theta}{\gamma}\left[\frac{1}{\hat{A}}\left[\ln(\hat{A}(e^{\gamma t}-1)+\hat{B}) - \ln(\hat{B})\right] - \frac{1}{\hat{A}-\hat{B}}\left[\ln(\hat{A}(e^{\gamma t}-1)+\hat{B}) - \ln(\hat{B})\right] - \frac{1}{\hat{B}-\hat{A}}\ln(e^{\gamma t})\right],\\
&= -\frac{2\kappa\theta}{\gamma}\left[-\frac{1}{\hat{A}}\left[\ln(\hat{B}) - \ln(\hat{A}(e^{\gamma t}-1)+\hat{B})\right] + \frac{1}{\hat{A}-\hat{B}}\left[\ln(\hat{B}) - \ln(\hat{A}(e^{\gamma t}-1)+\hat{B})\right] + \frac{1}{\hat{A}-\hat{B}}\ln(e^{\gamma t})\right],\\
&= -\frac{2\kappa\theta}{\gamma}\left[-\frac{1}{\hat{A}}\ln\left(\frac{\hat{B}}{\hat{A}(e^{\gamma t}-1)+\hat{B}}\right) + \frac{1}{(\hat{A}-\hat{B})}\ln\left(\frac{\hat{B}}{\hat{A}(e^{\gamma t}-1)+\hat{B}}\right) + \frac{1}{\hat{A}-\hat{B}}\ln(e^{\gamma t})\right],\\
&= -\frac{2\kappa\theta}{\gamma}\left[\ln\left[\left(\frac{\hat{B}}{\hat{A}(e^{\gamma t}-1)+\hat{B}}\right)^{-\frac{1}{\hat{A}}+\frac{1}{(\hat{A}-\hat{B})}}\right] + \frac{1}{\hat{A}-\hat{B}}\ln(e^{\gamma t})\right],\\
&= -\frac{2\kappa\theta}{\gamma}\left[\ln\left[\left(\frac{\hat{B}}{\hat{A}(e^{\gamma t}-1)+\hat{B}}\right)^{\frac{-(\hat{A}-\hat{B})+\hat{A}}{\hat{A}(\hat{A}-\hat{B})}}\right] + \ln\left[(e^{\gamma t})^{\frac{1}{\hat{A}-\hat{B}}}\right]\right],\\
&= -\frac{2\kappa\theta}{\gamma}\left[\ln\left[\left(\frac{\hat{B}}{\hat{A}(e^{\gamma t}-1)+\hat{B}}\right)^{\frac{\hat{B}}{\hat{A}(\hat{A}-\hat{B})}}\right] + \ln\left[\left((e^{\gamma t})^{\frac{\hat{A}}{\hat{B}}}\right)^{\frac{\hat{B}}{\hat{A}(\hat{A}-\hat{B})}}\right]\right],\\
&= -\frac{2\kappa\theta}{\gamma}\left(\ln\left[\left(\frac{\hat{B}e^{\frac{\hat{A}\gamma t}{\hat{B}}}}{\hat{A}(e^{\gamma t}-1)+\hat{B}}\right)^{\frac{\hat{B}}{\hat{A}(\hat{A}-\hat{B})}}\right]\right),\\
&= \ln\left[\left(\frac{\hat{B}e^{\frac{\hat{A}\gamma t}{\hat{B}}}}{\hat{A}(e^{\gamma t}-1)+\hat{B}}\right)^{\frac{-2\hat{B}\kappa\theta}{\gamma\hat{A}(\hat{A}-\hat{B})}}\right].
\end{aligned}$$

All that remains is to resubstitute our initial expressions for $\hat{A}$ and $\hat{B}$ in equation (91) back into our equation (94) and simplify. Thus, we have the first exponent as

$$\begin{aligned}\frac{-2\hat{B}\kappa\theta}{\gamma\hat{A}(\hat{A}-\hat{B})} &= \frac{-(2\gamma)(2\kappa\theta)}{\gamma(\gamma+\kappa+\lambda)((\gamma+\kappa+\lambda)-2\gamma)}, \\ &= \frac{-4\kappa\theta}{((\kappa+\lambda)+\gamma)((\kappa+\lambda)-\gamma)}, \\ &= \frac{-4\kappa\theta}{(\kappa+\lambda)^2-\gamma^2}, \\ &= \frac{-4\kappa\theta}{(\kappa+\lambda)^2-(\sqrt{(\kappa+\lambda)^2+2\sigma^2})^2}, \\ &= \frac{2\kappa\theta}{\sigma^2}.\end{aligned} \tag{95}$$

Similarly, the second exponent has the form

$$\begin{aligned}\frac{\hat{A}\gamma t}{\hat{B}} &= \frac{(\gamma+\kappa+\lambda)\gamma t}{2\gamma}, \\ &= \frac{(\kappa+\lambda+\gamma)t}{2}.\end{aligned} \tag{96}$$

Plugging equations (95) and (96) back into the hard-earned final expression in equation (94), we have the final desired solution:

$$A(\tau) = \ln\left(\frac{2\gamma e^{\frac{(\kappa+\lambda+\gamma)t}{2}}}{(\gamma+\kappa+\lambda)(e^{\gamma t}-1)+2\gamma}\right)^{\frac{2\kappa\theta}{\sigma^2}}. \tag{97}$$

Appendix B: Solving the SDE

While a variety of alternatives exist for discretizing the stochastic differential equations that we are dealing with, in the case of the Vasicek model we can actually solve for the diffusion process explicitly. To see how this is done, consider the following diffusion process postulated in the Vasicek model:

$$dr(t) = \kappa(\theta - r(t))dt + \sigma dW(t). \tag{98}$$

We now strip out the drift term and denote it as

$$Y(t) = \kappa(\theta - r(t)). \tag{99}$$

The next step is a bit odd, but necessary, and involves pre-multiplying equation (99) by $e^{\kappa t}$. This operation yields

$$e^{\kappa t}Y(t) = f(r(t)) = e^{\kappa t}\left(\kappa(\theta - r(t))\right). \tag{100}$$

This is, in fact, the trick in the derivation. We have a function that depends on a stochastic process, $r(t)$. We would like to describe its differential dynamics, and to do so we may apply Itô's theorem. Before we do this, let us compute the required partial derivatives:

$$\frac{\partial f(r(t))}{\partial t} = \kappa e^{\kappa t}Y(t), \tag{101}$$

$$\frac{\partial f(r(t))}{\partial r} = -\kappa e^{\kappa t}, \tag{102}$$

$$\frac{\partial f^2(r(t))}{\partial r^2} = 0. \tag{103}$$

We now have everything that we need to apply Itô's theorem. In particular,

$$\begin{aligned} f(r(t)) - f(r(0)) &= \int_0^t \underbrace{\frac{\partial f(r(s))}{\partial s}}_{\text{equation (101)}} ds + \int_0^t \underbrace{\frac{\partial f(r(s))}{\partial r}}_{\text{equation (102)}} dr(s) + \frac{1}{2}\int_0^t \underbrace{\frac{\partial f^2(r(s))}{\partial r^2}}_{\text{equation (103)}} d\langle r\rangle(s), \\ &= \int_0^t \kappa e^{\kappa s}Y(s)ds - \int_0^t \kappa e^{\kappa s}\left(\underbrace{Y(s)ds + \sigma dW(s)}_{\text{equation (98)}}\right), \\ &= -\int_0^t \kappa e^{\kappa s}\sigma dW(s). \end{aligned} \tag{104}$$

Inspection of equation (104) reveals that we have a recursive expression for $r(t)$ in terms of its previous value, $r(0)$. A bit of manipulation will make this clearer:

$$e^{\kappa t}Y(t) - e^{\kappa 0}Y(0) = -\int_0^t \kappa e^{\kappa s}\sigma dW(s), \tag{105}$$

$$\kappa e^{\kappa t}(\theta - r(t)) - \kappa(\theta - r(0)) = -\int_0^t \kappa e^{\kappa s}\sigma dW(s),$$

$$-\kappa e^{\kappa t}r(t) = -\kappa\theta e^{\kappa t} + \kappa\theta - \kappa r(0) - \int_0^t \kappa e^{\kappa s}\sigma dW(s),$$

$$r(t) = \underbrace{\theta\left(1 - e^{-\kappa t}\right) + e^{-\kappa t}r(0)}_{\text{Drift}} + \int_0^t \underbrace{e^{-\kappa(t-s)}\sigma dW(s)}_{\text{Diffusion}}.$$

In a sense, we are finished, as we have a recursive expression for $r(t)$ in terms of its previous value. Firstly, as in section 3, we evenly subdivide the interval $[0, T]$ into N subinterval and let $t_i = i\frac{T}{N}$ for $i = 1, ..., n$. In addition, we denote each time-step as $\Delta t = t_i - t_{i-1}$. In general, we have

$$r(t_i) = \theta\left(1 - e^{-\kappa\Delta t}\right) + e^{-\kappa\Delta t}r(t_{i-1}) + \epsilon(t_i) \tag{106}$$

where

$$\epsilon(t_i) = \int_{t_{i-1}}^{t_i} e^{-\kappa(t_i - s)}\sigma dW(s).$$

In other words, we have the first two moments of the Gaussian transition density of $r(t)$. Specifically,

$$r(t_i)|\mathcal{F}_{t_{i-1}} \sim \mathcal{N}\left(\underbrace{\theta\left(1 - e^{-\kappa\Delta t}\right) + e^{-\kappa\Delta t}r(t_{i-1})}_{\text{Mean}}, \underbrace{\epsilon^2(t_i)}_{\text{Variance}}\right). \tag{107}$$

All that remains, to get this expression into a form that can aid us in our simulations, is to find a more convenient way to express $\epsilon(t_i)$. In fact, it might not yet be obvious that $\epsilon(t_i)$ is actually the variance of our transition density. This is true by virtue of the fact that $\epsilon(t)$ is a stochastic integral and, as such, it has a zero expectation. We also recall that the quadratic variation process of the Brownian motion (i.e., $\langle W\rangle(t)$) is t. In particular, this means that

$$\mathbf{E}\left[\epsilon(t_i)|\,\mathcal{F}_{t_{i-1}}\right] = 0 \tag{108}$$

and that

$$\begin{aligned}
\textbf{var}\left[\epsilon(t_i)\middle|\mathcal{F}_{t_{i-1}}\right] &= \mathbf{E}\left[\epsilon^2(t_i)\middle|\mathcal{F}_{t_{i-1}}\right], \\
&= \mathbf{E}\left[\left(\int_{t_{i-1}}^{t_i} e^{-\kappa(t_i-s)}\sigma dW(s)\right)^2\middle|\mathcal{F}_{t_{i-1}}\right], \\
&= \mathbf{E}\left[\int_{t_i-1}^{t} e^{-2\kappa(t_i-s)}\sigma^2 ds\middle|\mathcal{F}_{t_{i-1}}\right], \\
&= \sigma^2\int_{t_{i-1}}^{t_i} e^{-2\kappa(t_i-s)}ds, \\
&= \frac{\sigma^2}{2\kappa}\left(1-e^{-2\kappa(t_i-t_{i-1})}\right), \\
&= \frac{\sigma^2}{2\kappa}\left(1-e^{-2\kappa\Delta t}\right).
\end{aligned} \tag{109}$$

As previously mentioned, the Vasicek transition density is actually known to be Gaussian and thus we may unabashedly state that

$$r(t_i)|\mathcal{F}_{t_{i-1}} \sim \mathcal{N}\left(\theta\left(1-e^{-\kappa\Delta t}\right)+e^{-\kappa\Delta t}r(t_{i-1}), \frac{\sigma^2}{2\kappa}\left(1-e^{-2\kappa\Delta t}\right)\right). \tag{110}$$

With the CIR model, however, the transition density follows a non-central χ^2-squared distribution, which is rather difficult to handle. Fortunately, Ball and Torous (1996) show that, over small time intervals, diffusions arising from stochastic differential equations behave like Brownian motion and, thus, to assume a normal transition density is probably a good approximation. Thus, for the purposes of simulation, we use the first two moments of the non-central χ^2-squared distribution and assume that

$$r(t_i)|\mathcal{F}_{t_{i-1}} \sim \mathcal{N}\left(\theta\left(1-e^{-\kappa\Delta t}\right)+e^{-\kappa\Delta t}r(t_{i-1}), \frac{\theta\sigma^2}{2\kappa}\left(1-e^{-\kappa\Delta t}\right)^2+\frac{\sigma^2}{\kappa}\left(e^{-\kappa\Delta t}-e^{-2\kappa\Delta t}\right)r(t_{i-1})\right). \tag{111}$$

Appendix C: Discrete-Time Affine Models

This appendix focuses on the derivation of these models in discrete time. This is not only an interesting exercise, but it also serves to broaden our understanding of these models and make accessible various other aspects of the literature. Nevertheless, the primary focus of the paper is the class of continuous-time affine term-structure models. This discussion appears, therefore, for completeness. Note that the approach is somewhat similar to, although relatively less mathematically sophisticated than, the continuous-time formulations.[39] The models also tend to be more computationally intensive than continuous-time models.

The foundation of the discrete-time approach comes from the following identity:[40]

$$1 = \mathbf{E}\left((1 + R_{n,t+1}) M_{t+1} | \mathcal{F}_t \right), \tag{112}$$

where n is the number of periods left to maturity of the underlying asset and t is the current time. What does this mean? First of all, M_{t+1} is termed the stochastic discount factor or the pricing kernel. Typically, in a utility maximization setting, the pricing kernel is interpreted as the intertemporal marginal rate of substitution. Second, $R_{n,t+1}$ is the real rate of return on some arbitrary asset over the period t to $t+1$. The meaning of this expression, therefore, is that the rate of return on any asset will depend on its covariance with the intertemporal marginal rate of substitution. In a fixed-income setting, however, the cash-flows are deterministic, which implies that the covariance of the cash-flows with the pricing kernel depends only on time. In this sense, we are developing a time series model for the pricing kernel.

Keeping the previous concept in the back of our mind, we now turn to look at the fixed-income market in particular. We have the following relationship, which represents the holding period return on a bond:

$$1 + R_{n,t+1} = \frac{P_{n-1,t+1}}{P_{n,t}}, \tag{113}$$

where P represents, as before, the value of a pure discount bond and n is the number of periods left to maturity of the pure discount bond and t. Now, we substitute equation (113) into (112),

$$1 = \mathbf{E}\left(\frac{P_{n-1,t+1}}{P_{n,t}} M_{t+1} \middle| \mathcal{F}_t \right), \tag{114}$$

and observe that $P_{n,t}$ is $\mathcal{F}_t$-measurable and thus constant:

$$P_{n,t} = \mathbf{E}\left(P_{n-1,t+1} M_{t+1} | \mathcal{F}_t \right). \tag{115}$$

[39]The majority of this discussion originates from Backus, Foresi, and Telmer (1998), Backus, Telmer, and Wu (1999), and Campbell, Lo, and MacKinlay (1997).

[40]Note that, as usual, we are defining these processes on a probability space, $(\Omega, \mathcal{F}, \mathbf{P})$. Moreover, we can consider $\mathcal{F}_t$ to be the natural filtration generated by M_{t+1}. This filtration is defined in the usual manner, as follows:

$$\mathcal{F}_t \triangleq \sigma\{M_s, s = 0, .., t\}.$$

This is a critical and intuitive result. It holds that the current price of a zero-coupon bond is the expectation, conditioned on all current information, of the product of the future price and the stochastic discount factor. This result also has an alternative representation that is sometimes useful. We know that $P_{0,t} = 1$ for all t, because the price of a dollar today is one dollar. Thus, we can use this to iterate on equation (114) as follows:

$$P_{1,t} = \mathbf{E}\left(\underbrace{P_{0,t+1}}_{P_{0,t+1}=1} M_{t+1} \middle| \mathcal{F}_t \right), \tag{116}$$

$$P_{1,t} = \mathbf{E}\left(M_{t+1} | \mathcal{F}_t \right).$$

Armed with the results in equations (114) and (116), which are the foundation of the entire discrete-time approach, let us move on to consider the term structure.[41] Our goal here is to highlight the similarities and understand the differences with the continuous-time approach. The first task is to describe the dynamics of our state variable, which we will call z. To this end, we begin with a continuous-time process and discretize it:

$$dz(t) = (b - az(t))dt + \sigma dW(t),$$

$$z_{t+1} - z_t = (b - az_t)\Delta t + \sigma\varepsilon_{t+1}.$$

At this point, we introduce a convenient normalization by requiring that $\Delta t = 1$,

$$z_{t+1} - z_t = a\left(\frac{b}{a} - z_t\right) + \sigma\varepsilon_{t+1}.$$

Now let

$$a = 1 - \varphi,$$

$$\frac{b}{a} = \theta.$$

[41] Equation (115) can be generalized in a relatively straightforward manner as

$$P_{n,t} = \mathbf{E}\left(\prod_{i=1}^{n} M_{t+i} \middle| \mathcal{F}_t \right). \tag{117}$$

A key step in this analysis uses a property of conditional expectation that is often called the *law of iterated expectations*. It holds that, given two σ-algebras $\mathcal{G} \subset \mathcal{F}$ and a random variable X defined on some probability space, $(\Omega, \mathcal{F}, \mathbb{P})$, the following two identities hold:

$$\mathbf{E}[\mathbf{E}[X|\mathcal{F}]|\mathcal{G}] = \mathbf{E}[X|\mathcal{G}],$$

$$\mathbf{E}[\mathbf{E}[X|\mathcal{G}]|\mathcal{F}] = \mathbf{E}[X|\mathcal{G}].$$

This rule is often quoted as "the smaller σ-algebra always wins." For more detailed discussion, see Durrett (1996).

This leads us to

$$z_{t+1} = z_t + (1-\varphi)(\theta - z_t) + \sigma\varepsilon_{t+1} \tag{118}$$
$$z_{t+1} = (1-\varphi)\theta + \varphi z_t + \sigma\varepsilon_{t+1}.$$

This is generally how the state variable is presented in discrete time. Note, however, that this representation follows from what is called the *Euler discretization*. To complete the model, we need to specify a process for the pricing kernel. It is more convenient to use the negative natural logarithm of the pricing kernel. This is because we will assume that the pricing kernel is conditionally log-normally distributed,

$$-\ln M_{t+1} = z_t + \beta\varepsilon_{t+1}, \tag{119}$$

where

$$\varepsilon_{t+1}|\varepsilon_t \sim \mathcal{N}(0,\sigma^2), \tag{120}$$

and, thus

$$\ln M_{t+1} \sim \mathcal{N}\left(-z_t, \beta^2\sigma^2\right). \tag{121}$$

It is hard to provide any intuition about this assumption other than offering the pragmatic fact that it contributes to greater tractability in the analysis. At this point, it will be useful to recall the following relationship about log-normal random variables. If $\ln X \sim \mathcal{N}(\mu,\sigma^2)$, then

$$\ln \mathbf{E}X = \mu + \frac{\sigma^2}{2}. \tag{122}$$

Applying this relationship to equation (116) yields the following expression:

$$P_{1,t} = \underbrace{\mathbf{E}(M_{t+1}|\mathcal{F}_t)}_{\text{equation (116)}}, \tag{123}$$
$$\ln P_{1,t} = \ln \mathbf{E}(M_{t+1}|\mathcal{F}_t),$$
$$\ln P_{1,t} = \underbrace{-z_t + \frac{\beta^2\sigma^2}{2}}_{\substack{\text{By equations} \\ \text{(121) and (122)}}}.$$

At this point, we postulate a general solution where we assume that the log bond price is an affine function of the state variable, z_t:

$$-\ln P_{n,t} = A_n + B_n z_t. \tag{124}$$

To find the form of A_n and B_n, we need to consider the conditional moments of $\ln M_{t+1} + \ln P_{n,t+1}$. To do this, we need to explicitly consider its form,

$$\begin{aligned}\ln(M_{t+1}P_{n,t+1}) = \ln M_{t+1} + \ln P_{n,t+1} &= \underbrace{-(z_t + \beta\varepsilon_{t+1})}_{\text{equation (119)}} - \underbrace{(A_n + B_n z_{t+1})}_{\text{equation (124)}}, \\ &= -z_t - \beta\varepsilon_{t+1} - A_n - B_n(\underbrace{(1-\varphi)\theta + \varphi z_t + \sigma\varepsilon_{t+1}}_{\text{equation (118)}}), \\ &= -(A_n + B_n(1-\varphi)\theta) - (1 + B_n\varphi)z_t - (\beta + B_n\sigma)\varepsilon_{t+1}.\end{aligned}$$

From this, we have the conditional moments,

$$\mathbf{E}(\ln M_{t+1} + \ln P_{n,t+1}|\mathcal{F}_t) = -(A_n + B_n(1-\varphi)\theta) - (1 + B_n\varphi)z_t, \tag{125}$$

$$\mathbf{var}(\ln M_{t+1} + \ln P_{n,t+1}|\mathcal{F}_t) = (\beta + B_n\sigma)^2\sigma^2, \tag{126}$$

or, rather,

$$\ln(P_{n,t+1}M_{t+1}) \mid \mathcal{F}_t \sim \mathcal{N}\left(-(A_n + B_n(1-\varphi)\theta) - (1 + B_n\varphi)z_t, (\beta + B_n\sigma)^2\sigma^2\right).$$

We now have all the required ingredients to determine the recursion relation for the general bond price. Again, by invoking the assumption of the log-normality of the pricing kernel, we have,

$$-\ln P_{n+1,t} = -\ln\left(\underbrace{\mathbf{E}(P_{n,t+1}M_{t+1}|\mathcal{F}_t)}_{\text{equation (115)}}\right) \tag{127}$$

$$\begin{aligned}&= -\mathbf{E}(\ln(M_{t+1}P_{n,t+1})|\mathcal{F}_t) - \frac{\mathbf{var}(\ln(M_{t+1}P_{n,t+1})|\mathcal{F}_t)}{2}, \\ &\qquad\qquad \overbrace{\qquad\qquad}^{\text{equation (126)}} \\ &= -\underbrace{\mathbf{E}(\ln M_{t+1} + \ln P_{n,t+1}|\mathcal{F}_t)}_{\text{equation (125)}} - \frac{\overbrace{\mathbf{var}(\ln M_{t+1} + \ln P_{n,t+1}|\mathcal{F}_t)}}{2}, \\ &= A_n + B_n(1-\varphi)\theta + (1 + B_n\varphi)z_t - \frac{(\beta + B_n\sigma)^2\sigma^2}{2}, \\ &= A_n + B_n(1-\varphi)\theta - \frac{(\beta + B_n\sigma)^2\sigma^2}{2} + (1 + B_n\varphi)z_t \\ &= A_{n+1} + B_{n+1}z_t,\end{aligned}$$

where

$$\begin{aligned}A_{n+1} &= A_n + B_n(1-\varphi)\theta - \frac{(\beta + B_n\sigma)^2\sigma^2}{2}, \\ B_{n+1} &= (1 + B_n\varphi).\end{aligned}$$

This is the discrete-time version of the closed-form solution for the bond price in the single-factor Vasicek model. Note that we have a formulation that is very similar to that which we derived in the continuous-time setting. Starting with a postulated set of stochastic processes for our underlying state variable, we used no-arbitrage arguments to construct a set of bond prices associated with the given value of the state variable. In this case, however, we have a recursion relation that represents the affine relationship between bond prices and the state variable.[42] This is in contrast to the function which we derived in the continuous-time world.

As before, we turn to the CIR model, which, although very similar in development, is slightly different, in that we permit the conditional variance to change over time. This is in contrast to the assumption of homoscedasticity that is imposed in the Vasicek model. To get to the discrete-time version of the single-factor CIR model, we merely postulate a new process for the state variable and the pricing kernel:

$$z_{t+1} = (1-\varphi)\theta + \varphi z_t + \sigma\sqrt{z_t}\varepsilon_{t+1}, \tag{128}$$

$$-\ln M_{t+1} = \left(1+\frac{\beta^2}{2}\right) z_t + \beta\sqrt{z_t}\varepsilon_{t+1}, \tag{129}$$

$$-\ln P_{n,t} = A_n + B_n z_t. \tag{130}$$

We now follow the same approach as in the previous section. That is, we calculate the conditional moments of our general expression:

$$\begin{aligned}
\ln M_{t+1} + \ln P_{n,t+1} &= \underbrace{-\left(1+\frac{\beta^2}{2}\right) z_t - \beta\sqrt{z_t}\varepsilon_{t+1}}_{\text{equation (129)}} \underbrace{-(A_n + B_n z_{t+1})}_{\text{equation (130)}}, \\
&= -\left(1+\frac{\beta^2}{2}\right) z_t - \beta\sqrt{z_t}\varepsilon_{t+1} - A_n - B_n(\underbrace{(1-\varphi)\theta + \varphi z_t + \sigma\sqrt{z_t}\varepsilon_{t+1}}_{\text{equation (128)}}), \\
&= -(A_n + B_n(1-\varphi)\theta) - (1+\frac{\beta^2}{2} + B_n\varphi) z_t - (\beta + B_n\sigma)\sqrt{z_t}\varepsilon_{t+1}.
\end{aligned}$$

That is, the moments of $\ln M_{t+1} + \ln P_{n,t+1}$ have the following form:

$$\ln\left(P_{n,t+1}M_{t+1}\right) \mid \mathcal{F}_t \sim \mathcal{N}\left(-(A_n + B_n(1-\varphi)\theta) - (1+\frac{\beta^2}{2} + B_n\varphi) z_t, (\beta + B_n\sigma)^2\sigma^2 z_t\right).$$

We then exploit the conditional log-normality of the pricing kernel and use the logic of equation (127) to

[42] As a quick check on this formula, we use the boundary conditions $A(0) = B(0) = 0$ and equation (124) with $n = 1$. This yields $\ln P_{1,t} = A_1 + B_1 z_t$. Given that $A_1 = -\frac{\beta^2\sigma^2}{2}$ and $B_1 = 1$, we find that the value of $P_{1,t}$ agrees with equation (123).

derive the following expression:

$$\begin{aligned} -\ln P_{n+1,t} &= -\mathbf{E}(\ln M_{t+1} + \ln P_{n,t+1}|\mathcal{F}_t) - \frac{\mathbf{var}(\ln M_{t+1} + \ln P_{n,t+1}|\mathcal{F}_t)}{2}, \\ &= A_n + B_n(1-\varphi)\theta + (1 + \frac{\beta^2}{2} + B_n\varphi)z_t - \frac{(\beta + B_n\sigma)^2 z_t \sigma^2}{2}, \\ &= A_n + B_n(1-\varphi)\theta + \left[1 + \frac{\beta^2}{2} + B_n\varphi - \frac{(\beta + B_n\sigma)^2\sigma^2}{2}\right] z_t, \\ &= A_{n+1} + B_{n+1} z_t, \end{aligned} \tag{131}$$

where

$$\begin{aligned} A_{n+1} &= A_n + B_n(1-\varphi)\theta \\ B_{n+1} &= 1 + \frac{\beta^2}{2} + B_n\varphi - \frac{(\beta + B_n\sigma)^2\sigma^2}{2} \end{aligned}$$

This recursion relation establishes the link between our state variable, $z(t)$, and bond prices, similar to equations (28) and (29) described in continuous time in section 2.3.[43]

[43] As another quick check on this formula, we use the boundary conditions $A(0) = B(0) = 0$ and equation (124) with $n = 1$ again yielding $\ln P_{1,t} = A_1 + B_1 z_t$. Given that $A_1 = 0$ and $B_1 = 1 + \frac{\beta^2(1+\sigma^2)}{2}$, we find that the value of $P_{1,t}$ agrees with equation (123).

Bibliography

ABKEN, P. A. (1990): "Innovations in the Modeling of the Term Structure of Interest Rates," *Federal Reserve Bank of Atlanta Economic Review*, 10, 2–27.

BABBS, S. H., AND K. B. NOWMAN (1999): "Kalman Filtering of Generalized Vasicek Term Structure Models," *Journal of Financial and Quantitative Analysis*, pp. 115–130.

BACKUS, D., S. FORESI, AND C. TELMER (1998): "Discrete-Time Models of Bond Pricing," Graduate School of Industrial Administration, Carnegie Mellon University.

BACKUS, D., C. TELMER, AND L. WU (1999): "Design and Estimation of Affine Yield Models," Graduate School of Industrial Administration, Carnegie Mellon University.

BALDUZZI, P., S. R. DAS, AND S. FORESI (1995): "The Central Tendency: A Second Factor in Bond Yields," Leonard Stern School of Business, New York University.

BALL, C. A., AND W. N. TOROUS (1996): "Unit Roots and the Estimation of Interest Rate Dynamics," *Journal of Empirical Finance*, 3, 215–238.

BEAGLEHOLE, D. R., AND M. S. TENNEY (1991): "General Solutions of Some Interest Rate-Contingent Claim Pricing Equations," *The Journal of Fixed Income*, 5, 69–83.

BILLINGSLEY, P. (1995): *Probability and Measure*. Wiley, Third Avenue, New York, New York, third edn.

BJÖRK, T. (1998): *Arbitrage Theory in Continuous Time*. Oxford University Press, Great Clarendon Street, Oxford, first edn.

BOLDER, D., AND D. STRÉLISKI (1999): "*Yield Curve Modelling at the Bank of Canada*," Technical Report No. 84. Ottawa: Bank of Canada.

BRENNAN, M. J., AND E. S. SCHWARTZ (1979): "A Continuous Time Approach to the Pricing of Bonds," *Journal of Banking and Finance*, 3, 133–155.

CAMPBELL, J. Y., A. W. LO, AND A. C. MACKINLAY (1997): *The Econometrics of Financial Markets*. Princeton University Press, Princeton, New Jersey.

CANABARRO, E. (1995): "Where do One-Factor Interest Rate Models Fail?," *The Journal of Fixed Income*, 4, 31–52.

CHAN, K. C., G. A. KAROLYI, F. A. LONGSTAFF, AND A. B. SANDERS (1992): "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *The Journal of Finance*, XLVII, 1209–1227.

CHAPLIN, G. B. (1996): "A Review of Term Structure Models and their Applications," Institute of Insurance and Pension Research Working Paper, University of Waterloo.

CHAPLIN, G. B., AND K. P. SHARP (1993): "Analytic Solutions for Bonds and Bond Options under n Correlated Stochastic Processes," Institute of Insurance and Pension Research Working Paper, University of Waterloo.

CHEN, R.-R. (1995): "A Two-Factor Preference-Free Model for Interest Rate Sensitive Claims Equilibrium Model of the Term Structure of Interest Rates," *Journal of Futures Markets*, 15, 345–372.

CHEN, R.-R., AND L. SCOTT (1993): "Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates," *Journal of Fixed Income*, 3, 14–31.

COX, J. C., J. E. INGERSOLL, AND S. A. ROSS (1985a): "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53, 363–384.

——— (1985b): "A Theory of Term Structure of Interest Rates," *Econometrica*, 53, 385–407.

DE JONG, F. (1998): "Time-Series and Cross-Section Information in Affine Term Structure Models," Department of Financial Management, University of Amsterdam.

DUAN, J.-C., AND J.-G. SIMONATO (1995): "Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter," Discussion paper, Centre universitaire de recherche et analyze des organizations (CIRANO).

DUFFIE, D. (1996): *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton, New Jersey, second edn.

DUFFIE, D., AND R. KAN (1996): "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6, 379–406.

DURRETT, R. (1996): *Probability: Theory and Examples*. Duxbury Press, Belmont, California, second edn.

FUNG, B. S., S. MITNICK, AND E. REMOLONA (1999): "Uncovering Inflation Expectations and Risk Premiums from Internationally Integrated Financial Markets," Bank of Canada Working Paper No. 99-6.

GEYER, A. L. J., AND S. PICHLER (1998): "A State-Space Approach to Estimate and Test Multifactor Cox-Ingersoll-Ross Models of the Term Structure of Interest Rates," Department of Operations Research, University of Economics Vienna.

GONG, F. F., AND E. M. REMOLONA (1996): "A Three-Factor Econometric Model of the U.S. Term Structure," Discussion Paper 9619, Federal Reserve Bank of New York.

HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

HARVEY, A. C. (1990): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, first edn.

HEATH, D., R. JARROW, AND A. MORTON (1990): "Bond Pricing and the Term Structure of Interest Rates: A Discrete Time Approximation," *Journal of Financial and Quantitative Analysis*, 25, 419–440.

——— (1992): "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation," *Econometrica*, 60, 77–105.

HO, T. S. Y., AND S.-B. LEE (1986): "Term Structure Movements and Pricing Contingent Claims," *The Journal of Finance*, XLI, 1011–1029.

HONORÉ, P. (1995): "Maximum Likelihood Estimation of Non-Linear Continuous Term Structure Models," Discussion paper, The Aarhus School of Business.

HULL, J., AND A. WHITE (1990): "Pricing Interest-Rate Derivative Securities," *The Review of Financial Studies*, 3, 573–592.

——— (1993): "One-Factor Interest Rate Models and Valuation of Interest Rate Derivative Securities," *Journal of Financial and Quantitative Analysis*, 28, 235–254.

——— (1994a): "Numerical Procedures for Implementing Term Structure Models I: Single-Factor Models," *The Journal of Derivatives*, pp. 7–16.

——— (1994b): "Numerical Procedures for Implementing Term Structure Models II: Two-Factor Models," *The Journal of Derivatives*, pp. 37–48.

JUDGE, G. G., W. GRIFFITHS, R. C. HILL, H. LÜTKEPOHL, AND T.-C. LEE (1985): *The Theory and Practice of Econometrics*. John Wiley and Sons, New York, New York, second edn.

KARATZAS, I., AND S. E. SHREVE (1991): *Brownian Motion and Stochastic Calculus*. Springer-Verlag, Berlin, second edn.

KIM, C.-J., AND C. R. NELSON (1999): *State-Space Models with Regime Switching*. The MIT Press, Cambridge, Massachusetts.

LANGETIEG, T. C. (1980): "A Multivariate Model of the Term Structure," *The Journal of Finance*, XXXV, 71–97.

LITTERMAN, R., AND J. SCHEINKMAN (1991): "Common Factors Affecting Bond Returns," *The Journal of Fixed Income*, 1, 54–61.

LONGSTAFF, F. A., AND E. S. SCHWARTZ (1992a): "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *The Journal of Finance*, XLVII, 1259–1282.

——— (1992b): "A Two-Factor Interest Rate Model and Contingent Claims Evaluation," *The Journal of Fixed Income*, pp. 16–23.

LUND, J. (1997): "Econometric Analysis of Continuous-Time Arbitrage-Free Models of the Term Structure of Interest Rates," Department of Finance, The Aarhus School of Business.

MC MANUS, D. J., AND D. G. WATT (1999): "Estimating One-Factor Models of Short-Term Interest Rates," Bank of Canada Working Paper No. 99-18.

MUSIELA, M., AND M. RUTKOWSKI (1998): *Martingale Methods in Financial Modelling*. Springer-Verlag, Berlin, first edn.

NOWMAN, K. B. (1997): "Gaussian Estimation of Single-Factor Continuous Time Models of the Term Structure of Interest Rates," *The Journal of Finance*, LII, 1695–1706.

OKSENDAL, B. K. (1995): *Stochastic Differential Equations*. Springer Verlag, Berlin, fourth edn.

PEARSON, N. D., AND T.-S. SUN (1994): "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox-Ingersoll-Ross Model," *The Journal of Finance*, XLIX, 1279–1304.

PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING, AND B. P. FLANNERY (1992): *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Trumpington Street, Cambridge, second edn.

RICHARD, S. F. (1978): "An Arbitrage Model of the Term Structure of Interest Rates," *Journal of Financial Economics*, 6, 33–57.

ROGERS, L. (1995): "Which Model for Term-Structure of Interest Rates Should One Use?," *The IMA Volumes in Mathematics and its Applications*, 67, 93–115.

SCHAEFER, S. M., AND E. S. SCHWARTZ (1984): "A Two-Factor Model of the Term Structure: An Approximate Analytical Solution," *Journal of Financial and Quantitative Analysis*, 4, 413–424.

SCHÖGL, E., AND D. SOMMER (1997): "Factor Models and the Shape of the Term Structure," Department of Statistics, University of Bonn.

SINGH, M. K. (1995): "Estimation of Multifactor Cox, Ingersoll, and Ross Term Structure Model: Evidence on Volatility Structure and Parameter Stability," *Journal of Fixed Income*, 4, 8–28.

SUN, T.-S. (1992): "Real and Nominal Interest Rates: A Discrete-Time Model and Its Continuous Time Limit," *The Review of Financial Studies*, 5, 581–611.

TIAN, W., P. P. BOYLE, AND F. GUAN (2000): "New Closed Form Solutions for the Extended CIR Model," Centre for Advanced Studies in Finance, University of Waterloo.

VASICEK, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.

Bank of Canada Working Papers
Documents de travail de la Banque du Canada

Working papers are generally published in the language of the author, with an abstract in both official languages. ***Les documents de travail sont publiés généralement dans la langue utilisée par les auteurs; ils sont cependant précédés d'un résumé bilingue.***

2001

2001-14	L'effet de la richesse sur la consommation aux États-Unis	Y. Desnoyers
2001-13	Predetermined Prices and the Persistent Effects of Money on Output	M.B. Devereux and J. Yetman
2001-12	Evaluating Linear and Non-Linear Time-Varying Forecast-Combination Methods	F. Li and G. Tkacz
2001-11	Gaining Credibility for Inflation Targets	J. Yetman
2001-10	The Future Prospects for National Financial Markets and Trading Centres	C. Gaa, S. Lumpkin, R. Ogrodnik, and P. Thurlow
2001-9	Testing for a Structural Break in the Volatility of Real GDP Growth in Canada	A. Debs
2001-8	How Rigid Are Nominal-Wage Rates?	A. Crawford
2001-7	Downward Nominal-Wage Rigidity: Micro Evidence from Tobit Models	A. Crawford and G. Wright
2001-6	The Zero Bound on Nominal Interest Rates: How Important Is It?	D. Amirault and B. O'Reilly
2001-5	Reactions of Canadian Interest Rates to Macroeconomic Announcements: Implications for Monetary Policy Transparency	T. Gravelle and R. Moessner
2001-4	On the Nature and the Stability of the Canadian Phillips Curve	M. Kichian
2001-3	On Commodity-Sensitive Currencies and Inflation Targeting	K. Clinton
2001-2	Exact Non-Parametric Tests for a Random Walk with Unknown Drift under Conditional Heteroscedasticity	R. Luger
2001-1	The Elements of the Global Network for Large-Value Funds Transfers	J.F. Dingle

2000

2000-23	The Application of Artificial Neural Networks to Exchange Rate Forecasting: The Role of Market Microstructure Variables	N. Gradojevic and J. Yang
2000-22	Une analyse empirique du lien entre la productivité et le taux de change réel Canada-É-U	D. Dupuis et D. Tessier

Copies and a complete list of working papers are available from:
Pour obtenir des exemplaires et une liste complète des documents de travail, prière de s'adresser à :

Publications Distribution, Bank of Canada
234 Wellington Street, Ottawa, Ontario K1A 0G9
E-mail: publications@bankofcanada.ca
Web site: http://www.bankofcanada.ca

Diffusion des publications, Banque du Canada
234, rue Wellington, Ottawa (Ontario) K1A 0G9
Adresse électronique : publications@banqueducanada.ca
Site Web : http://www.banqueducanada.ca